



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

의학박사 학위논문

일관성 가정 검증이 불가능한  
개방형 네트워크 메타분석 결과의  
신뢰성 평가 방법 개발

Development of a method for assessing the  
credibility of conclusions from open-loop  
network meta-analysis under unverifiable  
consistency assumption

2020년 8월

서울대학교 대학원

협동과정 의료정보학 전공

윤 정 화

## 국 문 초 록

다수의 대안 의료 중재들 간 상대적인 효과크기 추정을 통해 비교효과를 평가하고자 할 때 주요 연구방법론으로 고려되는 체계적 문헌 고찰 기반의 네트워크 메타분석은, 여러 중재들에 대한 상호 무작위배정 비교임상시험(Randomized Controlled Trial, RCT)들을 연결하여 근거 네트워크를 형성하고 각 연구들로부터 추출한 데이터를 통합하는 통계적 분석 방법이다. 근거 네트워크는 네트워크 내에서 같은 비교를 다루는 RCT 간 동질성 가정, 분석에 포함된 모든 RCT들은 임상적 및 방법론적으로 유사하다는 전이성 가정, 중재 간 비교의 직접비교 결과와 간접비교 결과의 일관성 가정을 만족해야 한다. 그러나 대안치료들 간 직접비교 연구가 부재한 상황에서 구성되는 개방형 네트워크에서는 일관성에 대한 통계적인 검정이 실제로도 불가능함에도 많은 경우 일관성을 가정하고 네트워크 메타분석을 수행한다.

우리나라와 유사한 의료기술평가 체계를 가지고 있는 캐나다, 영국, 호주의 의료기술평가 지침에서는 모두 네트워크 메타분석을 수행할 때 네트워크 내 근거의 일관성에 대해 탐색하여 보고하도록 하고있다. 캐나다와 호주의 경우, 공통대조군들의 결과값을 활용하여 연구 간 동질성을 평가하고 측정하기 어려운 환자 특성의 대리 지표로서 이를 사용하도록 하였다. 특히 호주 지침의 경우 메타회귀분석 외에도 개별환자데이터를 가용할 수 있는 경우 이를

이용한 보정 방법 사용 등 비일관성의 원인이 될 수 있는 요인들을 정량적으로 파악하고 완화시킬 것을 안내하고 있다.

서로 간 상호 비교된 바 없었던 제2형 당뇨병 환자 대상의 인슐린 병합 치료제들인 DPP4i, GLP-1RA, SGLT2i, TZD 간의 비교효과 평가를 위해, 해당 약제들의 인슐린 병합 치료군을 인슐린 치료군과 비교하였던 무작위배정 비교임상시험 연구들을 체계적 문헌고찰을 통해 선정하고 50개의 RCT로 구성된 개방형 네트워크 메타분석을 수행하였으며, 본 분석 수행을 통해 잠재적 교란요인을 평가하고 이를 보정하였을 때의 결론이 보정 전의 결론과 차이가 있을 수 있음을 제시하였다. 특히, 공통대조군인 인슐린군에서 관찰되는 결과 값을 기저 위험률 지표로서의 역할을 밝히고 이를 통한 보정이 결론에 미칠 수 있는 영향을 파악하여 제시하였다. 이질성이 보이는 RCT들로 구성된 개방형 네트워크에 대하여 일관성 가정하에 근거를 통합하였을 때, 각 치료제들 간에 불분명하게 나타났던 치료효과 차이는 각 연구들에서 상이하였던 기저 위험 수준을 동일하게 보정함으로써 명확히 구분될 수 있었음을 확인 하였고, 해당 분석 과정을 통해 일관성 검증이 불가능한 개방형 네트워크에서 전이성의 조건을 최대한 확보하여 잠재적인 일관성 위배를 최소화하는 접근의 중요성을 제시하였다.

네트워크 메타분석 접근을 사용하여 출판된 기존 연구 문헌들 중 개방형 네트워크 메타분석의 빈도 및 해당 연구들의 일관성에 대한 접근 방법과 결론의 한계점 등에 대한 경험적 근거 파악을 위해

수행한 체계적 문헌 고찰을 통하여 약 1년간 출판된 113개의 개방형 네트워크 메타분석 문헌을 검토한 결과, 네트워크 메타회귀분석 등의 구체적인 보정 방법을 적용하여 분석 결과에 반영한 연구는 단 10개에 지나지 않음을 확인하였다. 나머지 연구 중 96개의 연구에서는 일관성에 대한 가정 평가를 수행하지 않았거나(33개), 연구 특성의 정성적 비교만을 수행하여 일관성을 가정한 경우(57개), 또는 정성적 비교를 통해 비일관성이 예상됨에도 자료 수 부족으로 교란요인 보정 방법의 고려가 불가능하였던 경우(6개)였으나, 모두 일관성 가정하에서 자료 통합을 수행하였다. 이처럼 출판된 대다수의 개방형 네트워크 메타분석들은 확신할 수 없는 일관성 가정하에 간접비교결과에만 기반하여 통합된 결과가 제시되고 있었음을 확인할 수 있었고, 본 정성적 체계적 문헌고찰을 통해 이들 연구에 실제로 직접비교 연구 자료가 포함될 수 있었다면 기존 결론이 유지될 수 있을 것인가에는 심각한 불확실성이 존재하며 이러한 불확실성이 최종적인 결론 및 의료 의사결정에 반영에 될 수 있어야 함에 대한 문제를 제기하였다.

최종적으로 근거 네트워크에 포함된 연구들 간 특성의 정성적 평가 만을 통해 일관성이 충족될 수 있을 것임을 가정하여 근거 통합이 수행되는 개방형 네트워크 메타분석에서 얻어지는 결과의 신뢰성 평가를 위해, 개방형 네트워크 가상의 직접비교 연구 자료를 생성하여 산입(算入) 후 통합하여 기존 결과를 보정하고

보정 후 기존 결론의 유지될 수 있을 것인지에 대한 확실성을 평가하는 ‘자료산입을 통한 보정법(adjustment through imputation)’을 개발하였다. 생성되는 가상의 직접 비교 자료는 주어진 간접비교 결과와 일정 크기의 비일관성이 존재하되 통계적으로 허용될 수 있는(일관성 가정이 통계적으로 기각되지 않는) 수준이 되도록 하였고, 직접비교 결과는 간접비교 결과보다 보수적인 효과크기를 갖도록 조건을 설정하였다. 제안한 알고리즘을 통해 비일관성 크기의 허용범위를 찾고, 허용 범위 내에서 비일관성 증가에 따른 자료산입을 통한 보정법을 적용한 결과를 도출한 후 기존 결과의 유지 정도를 평가하여 기존 결과의 신뢰성을 평가하였다. 예시 시뮬레이션을 통해 본 방법의 타당성을 확인하였고, 시나리오에 따라 생성된 가상의 개방형 네트워크들에 본 방법을 적용하고 결과를 비교하여 적용성을 확인하였다.

의료기술평가 과정에서 네트워크 메타분석을 통해 제시된 근거에 기반하여 의사결정을 이루기 위해 참고하도록 여러 나라의 의료기술평가기관에서 제시한 지침에서는, 네트워크 메타분석을 수행할 때 일관성에 대한 탐색 및 보고를 하도록 권고하고 있으며, 개방형 네트워크에 기반한 분석을 수행하여 결과를 제공할 때에는 교란요인을 탐색하고 보정을 통하여 일관성 확보를 최대한 소명하여야 한다. 그러나 의료기술평가 현장에서 수행되는 많은 비교효과연구에서는 교란요인 보정을 수행하기에는 분석에 포함된 자료수가 충분하지 않은 경우가 많으며 개별환자자료 확보는

현실적으로 더욱 어렵다. 실제로 많은 연구들은 정성적인 평가에 기반하여 포함된 연구들이 임상적으로 매우 유사하다고 판단하고 이를 통해 일관성은 가정될 수 있다고 전제하여 분석을 수행하고 있으나, 이 경우 ‘통계적으로 허용될 수는 있으나 결과의 불확실성으로써 분석에 반영되어야 할’ 일정 수준의 비일관성은 실제로 존재할 수밖에 없고 이는 사실상 효과추정 결과 및 추론에 반영되어야 함을 인지하여야 하며, 본 연구를 통해 개발된 방법론은 이와 같은 상황에 적용되어 결과의 신뢰성을 평가하고 보수적인 결론에 반영하기 위해 유용하게 활용될 수 있을 것이다.

**주요어 :** 네트워크 메타분석, 일관성 가정, 개방형 네트워크,  
비일관성, 민감도 분석, 자료산입법

**학 번 :** 2014-30677

# 목 차

국문초록 .....	i
제1장. 서론 .....	1
제2장. 개방형 네트워크 메타분석의 수행 .....	7
2.1. 서론 .....	7
2.2. 네트워크 메타분석의 수행 .....	8
2.2.1. 기본 가정 .....	8
2.2.2. 모형 .....	10
2.2.3. 일관성 가정 검정 방법 .....	12
2.3. 개방형 네트워크 메타분석 .....	14
2.3.1. 여러 나라의 지침 .....	15
2.3.2. 일관성 확보를 위한 교란요인 보정 .....	20
2.4. 고찰 및 결론 .....	28
제3장. 개방형 네트워크 메타분석에 대한 체계적 문헌 고찰 ..	
.....	30
3.1. 서론 .....	30
3.2. 방법 .....	31
3.2.1. 문헌 검색 및 선정 .....	31



3.2.2. 자료 추출 및 요약.....	32
3.3. 결과.....	34
3.4. 고찰 및 결론.....	41
제4장. 개방형 네트워크 메타분석 결과의 신뢰성 평가 방법 ..	
.....	44
4.1. 서론.....	44
4.2. 방법.....	46
4.2.1. 통계적 방법.....	46
4.2.1.1. 자료산입 전략.....	48
4.2.1.2. 자료산입을 통한 보정법.....	50
4.2.2. 데이터 적용.....	53
4.2.2.1. Smoking cessation data set.....	53
4.2.2.2. Simulated data sets.....	56
4.3. 결과.....	59
4.3.1. Smoking cessation data set.....	59
4.3.2. Simulated data sets.....	66
4.4. 고찰 및 결론.....	70
제5장. 결론.....	77
참고문헌.....	81

부록 A.....	91
A.1. Summary of the studies included in the network meta-analysis.....	91
A.2. Weighted mean change in HbA1c from baseline in control group.....	97
A.3. Pairwise results of comparisons between antidiabetic agents as an add-on to pre-existing insulin therapy from network meta-analyses adjusted by study-level covariates .....	99
부록 B.....	100
B.1. $\widehat{var}(\hat{\theta}_{kBC}^*)$ , $k=1, \dots, l$ 를 계산하기 위한 공식 유도 과정 .....	100
B.2. R code : smoking cessation data set .....	101
B.3. R code : simulated data set (a) .....	112
B.4. Contrast-level forest plot with invariant intervals for the smoking cessation example .....	119
Abstract.....	120

## 표 목차

Table 1. 각국 의료기술평가 지침의 개방형 네트워크 메타분석 관련 내용 요약.....	19
Table 2. 자료 항목 .....	33
Table 3. 출판된 개방형 네트워크 특성 .....	36
Table 4. 가정 평가 수행 현황.....	38
Table 5. 분석 방법 적용 현황.....	40
Table 6. Simulated data sets .....	58
Table 7. Results of a network meta-analysis using the complete network and derived open-loop network for the smoking cessation data .....	60

## 그림 목차

Figure 1. 근거 네트워크의 구조 예시 .....	3
Figure 2. Evidence network for diabetes example .....	22
Figure 3. Meta-regression between the change in HbA1c in the control group and the treatment difference in the change of HbA1c.....	24
Figure 4. Meta-regression between potential confounders and the treatment difference in the change of HbA1c.....	26
Figure 5. 문헌 선정 과정 흐름도 .....	34
Figure 6. A graphical representation of an open-loop network consisting of one common comparator treatment ( $T_0$ ) and $p$ active treatments ( $T_1, \dots, T_p$ ). ..	47
Figure 7. Example networks for smoking cessation data set .....	54
Figure 8. Residual deviances by model type (y-axis) against the absolute extent of inconsistency (x-axis) .....	61
Figure 9. Interval plot of estimates of basic parameters	

against the extent of inconsistency (x-axis) within the obtained range.....	62
Figure 10. (a) Probability that each treatment is the best for smoking cessation against the extent of inconsistency within the obtained range. (b) Ranking of each treatment for successful smoking cessation against the extent of inconsistency within the obtained range.....	
	64
Figure 11. Residual deviances by model type (y-axis) against the absolute extent of inconsistency (x-axis) for each simulated data set. ....	
	67
Figure 12. Probability of each group being the best intervention (y-axis) against the extent of inconsistency, $ \omega_{BC} $ , (x-axis) within the obtained range for each data set. ....	
	68

# 제1장. 서론

임상 치료 가이드라인 및 정책에 대한 근거를 마련하기 위하여 시행되는 의료기술평가에서 관심 두 군간 상대효과를 평가하기 위해, 무작위배정임상시험(Randomized Controlled Trials, RCT)에 대한 체계적 문헌 고찰이 고려되고 RCT로부터 얻은 상대효과 결과값들을 통합하는 메타분석이 사용된다[1]. 만일 관심 중재(intervention) 간 RCT가 없다면, 이들 간 상대효과 비교를 위하여 제3의 치료 대안을 참조군으로 하는 간접비교(indirect comparison) 방법을 고려한다[2]. 3개 이상의 여러 중재 간 비교를 해야 할 때는 RCT로부터 얻은 직접비교 결과와 간접비교 결과를 모두 이용하여 혼합비교(mixed comparison)를 수행한다. 네트워크 메타분석은 간접비교와 혼합비교를 포함하는 일반적인 형태의 메타분석이다[3]. 네트워크 메타분석은 세가지 이상의 중재들에 대한 여러 RCT들로 연결된 근거 네트워크를 구성하고 각 연구들로부터 추출된 자료를 통합하는 통계적 방법이며[4, 5], 분석에 포함된 중재들의 유효성 및 안전성에 대한 결과값에 대하여 상대효과 크기를 비교하여 순위를 매길 수 있다[6].

근거 네트워크는 관심 중재 간 RCT 자료로 구성된다[3]. 다양한 네트워크 형태를 Figure 1에 예시하였다. 네트워크를 구성하는 각 노드(node) A, B, C, D는 치료군을 나타내고 두 노드를 연결한 직선은 두 치료군 간 RCT가 존재함을 의미한다. (a) 네트워크는 중

재 A와 B, A와 C의 RCT가 존재하는 열린 루프(open loop) 모양이며, 공통으로 사용된 대조군(A)에 대한 상대적인 치료효과를 통해 B와 C 간 상대효과를 간접적으로 추정한다. (b)와 (c)는 여러 개의 열린 루프들로 구성되었으며, 각각 별(star) 모양과 사다리(ladder) 모양을 가진다. 이와 같이 열린 루프만으로 구성된 네트워크를 ‘개방형 네트워크(open-loop network)’라 한다[7]. 모든 중재 간 RCT가 존재하는 경우에는 (d) 네트워크와 같이 닫힌 루프(closed loop) 형태를 갖는다. (e)는 닫힌 루프와 열린 루프를 포함하는 네트워크이고 (f)는 모든 중재 간 RCT가 존재하여 닫힌 루프만으로 구성된 네트워크이다.

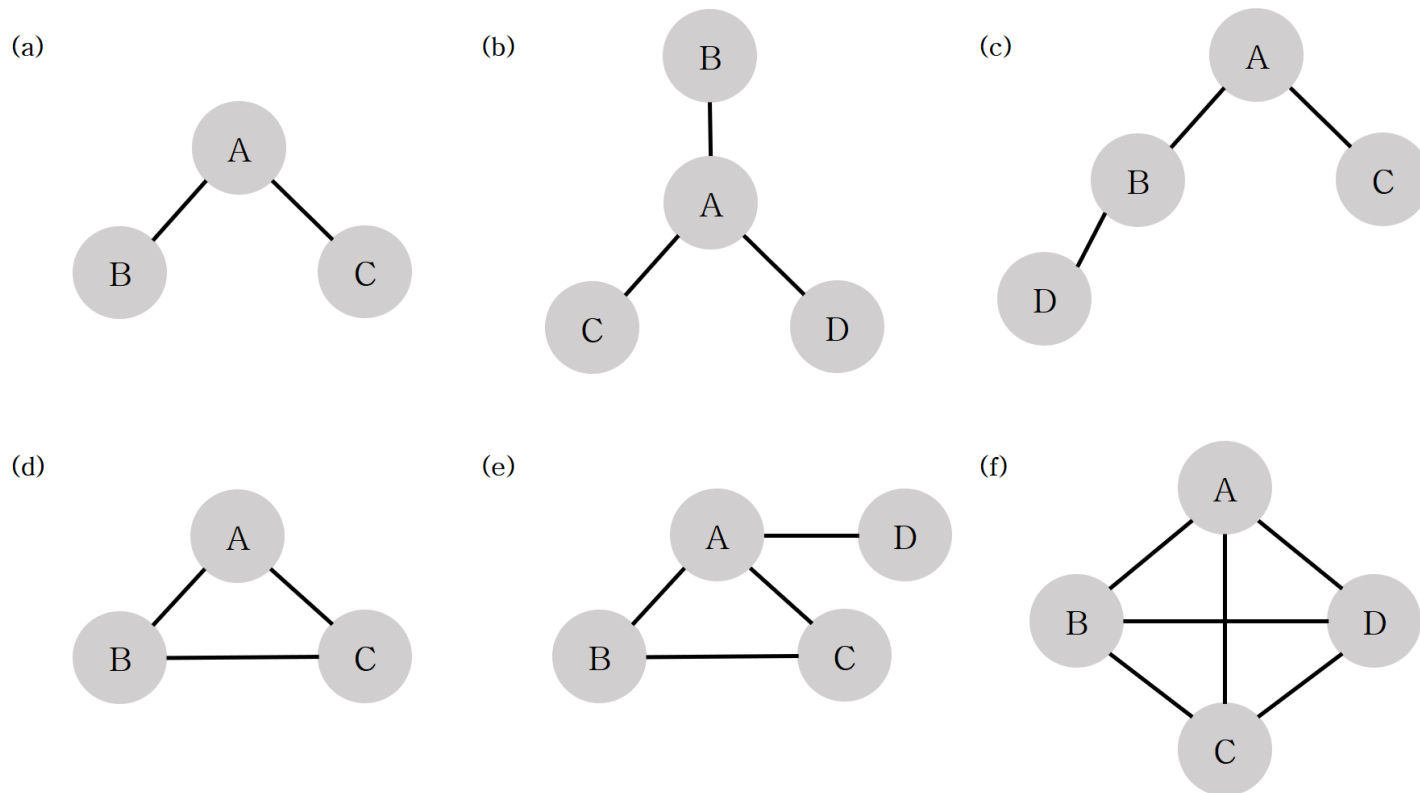


Figure 1. 근거 네트워크의 구조 예시



네트워크 메타분석은 치료군 간 직접비교 RCT를 통합하는 쌍별 비교 메타분석(pairwise meta-analysis)을 수행하는 동시에 직접 비교와 간접비교 근거를 통합하는 혼합비교를 수행한다. 그러므로 근거 네트워크는 같은 비교를 다루는 RCT들 간 동질성(homogeneity) 가정[8]을 만족하고, 네트워크에 포함된 모든 RCT들은 임상적 및 방법론적으로 유사하다는 전이성(transitivity) 가정을 만족해야 한다. 이를 교환가능성(exchangeability)이라고도 한다[9]. 또한, 네트워크 내 동일 비교에 대한 직접비교 결과값과 간접비교 결과값 간 일관성(consistency) 가정을 만족해야 한다[4, 10, 11]. 일관성을 검정하기 위해 직접비교 와 간접비교의 결과값 간 차이, 즉 비일관성(inconsistency)을 정의한다[12]. 닫힌 루프를 포함하는 근거 네트워크의 경우, 특정 닫힌 루프 또는 네트워크 전반에서 비일관성에 대한 통계적 유의미성을 평가한다. 그러나 열린 루프로만 구성된 개방형 네트워크의 경우 일관성 가정에 대한 검정이 불가능하다. 본 연구에서는 일관성 가정 검정이 불가능한 개방형 네트워크 중 별 모양 네트워크에 대한 네트워크 메타분석에 초점을 두었으며, 개방형 네트워크 메타분석의 적절성 및 한계점을 구체화하고 그 한계점을 극복하기 위한 대안을 연구하였다.

신약의 시판 허가 승인을 얻기 위해 위약이나 표준치료 비교연구를 수행하는 경우[13], 그리고 허가 승인을 받은 후 해당 약을 다른 활성 치료제들과 비교하지 않는 경우[14] 개방형 네트워크가 형성된다. Chaimani et al (2013)[15]에 따르면, PubMed에서 2011

년 3월까지 발표된 152개의 네트워크 메타분석 중 47개는 개방형 네트워크였다[15]. 개방형 네트워크에서는 일관성 가정 검정이 불가능하기 때문에 분석에 포함된 연구들 간 동질성 및 전이성 가정에 대한 검정을 수행하는 것이 중요하다[9, 16]. 일반적으로 동질성, 전이성, 일관성 가정의 위배에 영향을 미치는 요인(연구 및 환자 특성, 결과값의 정확도 등)은 비슷하기 때문에, 동질성 및 전이성 가정을 만족하는 개방형 네트워크의 경우 일관성 가정을 만족할 가능성이 높다. 하지만, 개방형 네트워크에서는 일관성 가정에 대한 식별이 불가능하다는 불확실성이 있으며, 이러한 불확실성은 분석에서 반영되지 않기 때문에 분석 결과의 신뢰성 확보에 한계가 있다. 적어도 한 개 이상의 닫힌 루프를 포함하는 네트워크의 경우, 랜덤 효과 모형이 고려될 때 통계적으로 허용되는 비일관성의 크기가 연구 간 이질성의 일부로 취급되어 통합 효과 크기 추정에 반영된다. 그러나 개방형 네트워크 메타분석에서는 이러한 분석적 고려가 불가능하다. 따라서 직접비교 연구가 부재한 개방형 네트워크에서의 일관성 가정에 대한 불확실성이 분석 결과에 어느 정도 영향을 미치는지 평가할 수 있는 방법론 개발이 요구되며, 이는 의료기술평가를 통한 의사결정과정에서 필수적으로 다루어져야 한다.

2장에서는 네트워크 메타분석의 기본가정과 네트워크 메타분석의 기본 모형을 구체적으로 설명하고, 개방형 네트워크 메타분석 수행과 관련하여 여러 나라의 의료기술평가 지침을 검토하고 정리하였다. 또한, 제2형 당뇨병 환자 대상의 인슐린 병합 치료제들 간의

비교효과 평가를 위한 개방형 네트워크 메타분석에서, 교란요인 보정을 통해 일관성 검증이 불가능한 개방형 네트워크에서 전이성을 최대한 확보하여 잠재적인 일관성 위배를 최소화하는 접근의 중요성을 제시하였다. 3장에서는 체계적 문헌 고찰을 통해 출판된 개방형 네트워크 메타분석들을 대상으로 네트워크의 특징, 가정 검정 및 분석 방법을 검토하여 개방형 네트워크 메타분석 수행 과정에 나타나는 실질적인 문제점 및 한계점을 고찰하였다. 4장에서는 확신할 수 없는 일관성 가정하에 수행된 개방형 네트워크 메타분석 결과의 신뢰성을 평가하기 위해 새롭게 개발된 방법론을 제안하였다.

## 제2장. 개방형 네트워크 메타분석의 수행

### 2.1. 서론

본 장에서는 네트워크 메타분석의 주요 가정인 동질성, 전이성, 일관성 가정을 상세하게 설명하였다. 또한, 네트워크 메타분석의 기본 모형을 설명하고, 개방형 네트워크에서 모형을 정의하였다. 우리나라와 유사한 의료기술평가 체계를 가지고 있는 캐나다, 영국, 호주 의 의료기술평가 자료제출 준비를 위한 지침에서 제안하는 개방형 네트워크 메타분석에 관한 내용을 검토하여 각국의 지침을 요약하고 비교하였다. 마지막으로 서로 상호 비교된 바 없는 제2형 당뇨병 환자에게 인슐린 병합요법으로 투약하는 치료제들, DPP4i, GLP-1RA, SGLT2i, TZD 간의 비교효과 평가를 위한 개방형 네트워크 메타분석에서 일관성 확보를 위한 교란요인 보정 과정 및 결과를 설명하였다.

## 2.2. 네트워크 메타분석의 수행

### 2.2.1. 기본 가정

네트워크 메타분석은 네트워크 내 동일 비교 연구들 간 동질성(homogeneity) 가정을 만족해야 한다[8]. 또한, 분석에 포함된 모든 연구들 간 전이성(transitivity) 가정을 만족해야 한다[9]. 전이성 가정을 교환가능성(exchangeability) 가정이라고 표현하기도 한다[17]. 전이성 가정은 치료 A와 B의 RCT와 치료 A와 C의 RCT에서 치료 B와 치료 C를 바꾸어 임상시험을 수행하더라도 기존의 임상시험에서 얻은 것과 동일한 효과를 얻을 수 있다는 가정이다. 전이성 가정을 평가하는 방법은 다음과 같다. 우선, 공통대조군(A)의 결과값 간 이질성을  $I^2$  혹은 Q 통계량을 이용하여 평가한다. 연구 및 환자 특성이 치료효과에 미치는 영향을 정성적으로 평가하거나 네트워크 메타회귀분석을 이용하여 정량적으로 평가한다[9, 18].

네트워크 메타분석은 일관성 가정을 만족해야 한다[11, 19]. 일관성 가정을 검정하기 위해 직접비교 치료효과와 간접비교 치료효과 간 차이, 즉, 비일관성(inconsistency)을 정의한다[12]. 이를 ‘루프 비일관성’이라 하며 일반적으로 비일관성은 루프 비일관성을 일컫는다. 여러 일관성 가정 검정 방법들은 비일관성을 계산하는 방법과 검정하는 방법에 따라 구분된다[2, 11, 12, 19–22]. 일관성 가정 평가 방법은 지역적(local) 혹은 광역적(global) 평가 방법으

로 구분할 수 있다. 지엽적 평가 방법은 근거 네트워크를 구성하는 모든 닫힌 루프에 대하여 개별적 및 독립적으로 일관성 가정을 평가하는 방법이다. 그 방법으로는 Bucher의 방법[2], Back calculation 방법[21], Node-splitting 방법[21] 등이 있다. Bucher의 방법은 단일 루프에서 간접비교 효과 크기를 계산하고 Back calculation 방법과 Node-splitting 방법은 특정 치료군 간 비교에 대한 직접비교 연구를 제외하고 네트워크 메타분석을 수행하여 해당 치료군 간 비교에 대한 간접비교 효과 크기를 계산하여, 직접비교 효과 크기와의 차이를 계산한다. 광역적 평가방법은 근거 네트워크를 구성하는 모든 데이터들을 기반으로 해당 네트워크의 일관성 가정을 총체적으로 평가하는 방법이며, 모형을 사용하여 평가한다. 모형의 종류에 따라 방법을 구분한다. 모형의 종류는 비일관성을 랜덤효과로 고려한 ‘변동 비일관성 베이지안 모형(random inconsistency Bayesian mixed treatment comparison model) [19]’, 루프 비일관성 뿐만 아니라 연구 설계에 따른 비일관성을 고려한 ‘치료별 연구 설계의 교호작용 효과 모형(design-by-treatment interaction model) [22, 23]’, 네트워크에 포함된 직접비교 RCT만으로 효과 크기를 추정하는 ‘관계없는 상대적 효과 평균을 통한 비일관성 모형(inconsistency model with unrelated mean relative effects) [12]’ 등이 있다.

### 2.2.2. 모형

체계적 문헌 고찰을 통해 선정된 RCT의 요약된 데이터를 이용한 네트워크 메타분석의 기본 모형은 다음과 같다.

$\hat{\theta}_{ijk}$ 는 study  $i$ 에서 관찰된 치료법  $j$ 에 대한 치료법  $k$ 의 상대적 효과 크기이고,  $\hat{\theta}_{ijk}$ 는 평균이  $\theta_{ijk}$ 이고, 분산이  $\sigma_{ijk}^2$ 인 정규분포를 따른다.  $\theta_{ijk}$ 는 study  $i$ 의 치료법  $j$ 에 대한 치료법  $k$ 의 study 치료효과이다.  $\widehat{var}(\hat{\theta}_{ijk})$ 은  $\hat{\theta}_{ijk}$ 의 추정된 분산으로 분산의 참값(true value)인  $\sigma_{ijk}^2$ 으로 고려하여, 위 분포는  $\hat{\theta}_{ijk} \sim N(\theta_{ijk}, \widehat{var}(\hat{\theta}_{ijk}))$ 을 만족한다고 가정한다.  $\theta_{ijk}$ 의 모형은 다음과 같다.

$$\theta_{ijk} \sim Normal(d_{jk}, \tau^2)$$

$d_{jk}$ 은 치료법  $k$ 에 대한 치료법  $j$ 의 평균 치료효과 크기이다. 네트워크 내 모든 비교들에 대해 공통의 연구 간 변동( $\tau^2$ )을 가정하였으며, 이는 분석의 간결성을 위한 가정으로 네트워크에 포함된 전체 비교 내 이질성 크기의 평균을 모수 추정에서 반영하는 것을 의미한다. 모형에서 각 비교 내 연구 간 변동은 별도로 정의될 수도 있다[10, 19].

네트워크 메타분석을 수행하는 표준 접근법에서, 우선, 위약 혹은 표준치료제를 기저 치료법(A)으로 선택하고, 이를 사용하여 기본 모수(basic parameter)  $d_{Ak}$  와  $d_{Aj}$  를 정의한다. 기능 모수

(functional parameter)  $d_{jk}$  ( $j$  and  $k \neq A$ )는 기본 모수를 사용하여 일관성 관계식  $d_{jk} = d_{Ak} - d_{Aj}$ 로 정의한다. 이러한 접근법을 기반으로 하는 모형은 일관성 모형 (consistency model)이라 한다.

치료법 A, B, C로 구성된 닫힌 루프인 가장 간단한 네트워크에 대하여 모형을 다음과 같이 표현할 수 있다.

$$\begin{aligned}\theta_{iAB} &\sim \text{Normal}(d_{AB}, \tau^2) \\ \theta_{iAC} &\sim \text{Normal}(d_{AC}, \tau^2) \\ \theta_{iBC} &\sim \text{Normal}(d_{AC} - d_{AB}, \tau^2) \quad *\end{aligned}$$

치료법 B와 C 간 상대적 치료효과 크기는 식  $\hat{d}_{BC} = \hat{d}_{AC} - \hat{d}_{AB}$ 을 통해 계산된다. 기본 모수 추정치를 바탕으로 치료 간 비교우위 (treatment ranking)를 선정 할 수 있다.



### 2.2.3. 일관성 가정 검정 방법

다양한 일관성 가정 검정 방법 중 ‘관계없는 상대적 효과 평균을 통한 비일관성 모형[12]’을 이용한 일관성 가정 검정 방법을 소개하였다. 본 모형은 일관성 가정 평가를 위해 사용되는 것이며, 네트워크에 포함된 각 치료법들 사이의 비교를 직접비교 RCT만으로 추정한다. 이를테면, 치료법 A, B, C로 구성된 닫힌 루프 네트워크에 대한 모형은 다음과 같이 표현된다.

$$\theta_{iAB} \sim \text{Normal}(d_{AB}, \tau^2)$$

$$\theta_{iAC} \sim \text{Normal}(d_{AC}, \tau^2)$$

$$\theta_{iBC} \sim \text{Normal}(d_{BC}, \tau^2)$$

기본 모수는  $d_{AB}$ ,  $d_{AC}$ ,  $d_{BC}$ 이다. 해당 모형을 통한 분석은 네트워크 내 비교들에 대한 쌍별 메타분석을 수행하는 것과 동일하다. 단, 연구 간 변동성을 공유하고 있을 뿐이다. B와 C의 RCT가 존재하지 않는 개방형 네트워크에서는 B와 C의 상대적 효과 크기를 추정할 수 없다.

위 비일관성 모형과 2.2.2장에서 소개한 일관성 모형을 이용하여 각모형의 residual deviance 값을 비교하여 일관성 가정을 검정한다[12]. Residual deviance 값이 더 작은 모형이 주어진 데이터셋을 잘 적합한 모형으로 고려된다. 따라서, 만일 비일관성 모형의

residual deviance 값이 더 작다면, 잠재적으로 일관성 가정을 충족하지 않는 것으로 간주한다.

## 2.3. 개방형 네트워크 메타분석

개방형 네트워크에서는 B와 C의 RCT가 존재하지 않기 때문에

2.2.2. 모형에서 \*이 제외된다. 모형은 다음과 같이 정의된다.

$$\theta_{iAB} \sim \text{Normal}(d_{AB}, \tau^2)$$

$$\theta_{iAC} \sim \text{Normal}(d_{AC}, \tau^2)$$

직접비교 RCT가 존재하지 않은 치료법 B와 C간 상대효과( $d_{BC}$ )는 일관성 가정 하에 기본 모수의 추정치  $\hat{d}_{AC}$ ,  $\hat{d}_{AB}$ 를 사용하여 일관성 가정하에 식  $\hat{d}_{BC} = \hat{d}_{AC} - \hat{d}_{AB}$ 을 통해 간접적으로 추정된다.

2.2.2의 치료법 A, B, C로 구성된 닫힌 루프에서는 기본 모수 ( $d_{AB}$ ,  $d_{AC}$ )의 추정에 치료법 B와 C 직접비교 RCT 데이터(\*)가 기여하고 있기 때문에 추정되는 기능 모수  $\hat{d}_{BC}$ 는 B와 C의 직접비교 결과와 간접비교 결과의 가중평균값으로 도출될 것이다. 그러나, 위 개방형 네트워크에서는 B와 C의 RCT가 존재하지 않기 때문에  $\hat{d}_{BC}$ 는 오직 간접비교만으로 결과가 도출된다.

### 2.3.1. 여러 나라의 지침

캐나다의 의료기술평가 지침 Guidelines for the Economic Evaluation of Health Technologies (CADTH, 2017) [24]에서는, 평가 목적에 적합한 신뢰할 수 있는 두 개 이상의 대조군 또는 직접 비교 된 바 없는 두 개 이상의 대조군을 포함하는 일반적인 네트워크 메타분석을 허용하며, 다음과 같은 분석 가이드를 제공하고 있다. 네트워크 메타분석에 포함된 자료들의 적합성, 즉, 네트워크 메타분석에 포함된 연구들이 분석의 목적과 범위 및 선정 기준과 제외 기준에 부합하여 분석을 통해 마련하고자 하는 근거를 기반으로 하는 의사결정과 연관되어 있는지 평가하는 것을 요구하고 있다. 통계 분석과 관련한 구체적인 방법론에 대해서는 Cooper et al. (2015) [25]을 참고하도록 제시하고 있다. 이 문헌에서는 네트워크 메타분석을 수행할 때 충족해야 하는 가정과 분석 방법을 다음과 같이 상세히 설명하였다. 우선, 쌍별 메타분석에서와 같이 네트워크 내 두 치료군 간 치료효과의 이질성을 램덤효과로 고려할 것 인지에 대하여 고정효과 모형과 랜덤효과 모형의 모형 적합도 통계량을 통해 결정하도록 하였다[25]. 추가적으로 네트워크를 구성하는 노드 간 연결되어 있는지, 직접비교 근거와 간접비교 근거의 일관성을 탐색하도록 하였다. 이질성과 비일관성은 치료효과와 상호작용을 일으키는 요인들로부터 기인할 수 있는데, 이 요인들의 개수는 상당히 많고 측정 불가능하거나 알 수 없는 경우도 있을 수 있으나, 만일 이 요

인들을 파악할 수 있고 측정할 수 있다면 이질성과 비일관성을 해소하기 위해 자료통합 모형에 반영할 수 있음을 제안하였다. 또한, 네트워크에 참조 치료군의 효과 크기는 측정할 수 없는 환자 특성의 대리 지표(proxy)로서 이질성 또는 비일관성의 잠재적 원인으로 고려될 수 있음을 언급하였다.

영국의 지침 Guide to the methods of technology appraisal(NICE, 2013) [26]에서는 직접비교 RCT 의 쌍별 메타분석을 기본분석으로 제시하는 것을 권고하고 있다. 하지만 RCT가 없는 경우, 공통대조군을 이용한 간접비교 및 혼합비교를 포함하는 네트워크 메타분석을 표준 쌍별 메타분석의 모범사례원칙을 준수하여 사용할 것을 제안하였다. 또한, 네트워크 메타분석에 포함된 개별 RCT의 연구 방법과 환자의 기저 특성을 상세히 기록하고 네트워크 메타분석 내 쌍별 메타분석에서 연구들의 이질성과 직접비교 근거와 간접비교 근거의 일관성에 대한 보고를 하도록 하였고 비일관성이 의심되면 그 원인을 소명하도록 하였다. 간접비교를 수행할 때 자료 합성 방법에 대한 상세한 설명과 RCT에 대한 검색과 선정 및 제외 기준에 대한 이론적 근거를 제시하고, 네트워크 메타분석에 대한 결과는 PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-analyses)-네트워크 메타분석 체크리스트[27]에 따라 보고하도록 하였으며, 이 체크리스트에서는 ‘네트워크 메타분석의 기본 가정에 대한 평가 수행’ 및 ‘가정이 충족되지 않을 때 이를 해결하기 위해 사용한 통계적 접근 방법’ 관련 기술을

요구하고 있다.

호주의 지침 Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC, 2016) [28]에서는 간접비교 분석에 포함된 RCT에 대한 대상자 기저 특성, 치료의 세부사항, 결과변수 정의 및 결과값을 제시하고, 이를 바탕으로 연구들의 동질성에 대해 정성적으로 평가하도록 하였다. 가능한 경우 잠재적 교란요인을 평가하는 네트워크 메타회귀분석을 수행하고, 공통대조군 결과값에 대한 연구 간 동질성에 대하여 정량적인 평가를 하도록 하였다. 또한, 직접비교 근거와 간접비교 근거의 일관성에 대한 보고를 하도록 하였다. 가정 충족이 불확실한 경우에는 교란요인을 보정한 결과를 최종 결과로 도출하도록 권고하였고, 보정 변수 한 개 당 최소 10개의 RCT가 존재하는 경우에만 보정 방법 활용이 타당하다고 명시하였다. 개별환자데이터를 사용할 수 있는 연구가 존재하는 경우에는 Matching-adjusted indirect comparison(MAIC) 와 Simulated treatment comparison(SCT)와 같은 Population-adjusted indirect comparisons방법들[29-31]을 적용하여 연구들 간 환자 특성의 불균형을 개선하도록 하였다.

국내 건강보험심사평가원에서 제시하는 경제성평가 지침(심평원, 2011) [32] 내에서는 등재신청약품과 비교대상약제의 비교 RCT자료가 없는 경우 위약이나 다른 제3의 치료대안을 참조대상으로 하여 등재신청의약품과 비교대상약제에 대한 간접비교 평가에 대하여 제시하고 있다. 그러나, 해당 지침에서는 네트워크 메타분석에 대해

구체적으로 언급하지 않았다. 또한, 간접비교 분석 수행과 관련한 자세한 내용은 별도로 마련된 간접비교 지침[33]을 참고하도록 하였다. 간접비교 지침에는 포함된 연구 간 전이성 가정을 충분히 소명하도록 하였으며, 해당 가정을 충분히 소명할 수 없는 경우 도출된 결과는 보조적 정보로 고려할 것으로 권고하였다.

위 각국 지침에서 제안하는 개방형 네트워크 메타분석관련 내용을 다음 표에 요약하였다(Table 1). 각국의 지침에서는 개방형 네트워크 메타분석 수행에 대해서 별도로 언급되어 있지 않았고 일반적인 네트워크 메타분석관련 지침을 안내하고 있었다. 캐나다, 영국, 호주 지침에서는 직접비교 근거와 간접비교 근거의 일관성에 대해 탐색하여 보고하도록 하였으며, 영국의 경우 일관성에 대한 불확실성이 큰 경우 그 원인을 소명하도록 하였다. 캐나다와 호주의 경우 공통대조군들의 결과값을 활용하여 연구 간 동질성을 평가하고 측정하기 힘든 환자 특성의 대리 지표로 사용하도록 하였다. 특히 호주 지침의 경우 메타회귀분석 또는 IPD를 이용한 Population-adjusted indirect comparisons방법을 사용하도록 하여 비일관성의 원인을 정량적으로 파악하고 완화시키도록 안내하고 있다.

Table 1. 각국 의료기술평가 지침의 개방형 네트워크 메타분석  
관련 내용 요약

구분	내용
캐나다 CADTH 2017	<ul style="list-style-type: none"> <li>- 네트워크 메타분석 수행 시 일관성 가정 탐색</li> <li>- 공통대조군의 결과값을 측정할 수 없는 환자 특성의 대리 지표로 활용</li> </ul>
영국 NICE 2013	<ul style="list-style-type: none"> <li>- 네트워크 메타분석 수행 시 일관성에 대한 보고</li> <li>- 일관성에 대한 불확실성이 클 때 관련 원인 탐색</li> </ul>
호주 PBAC 2016	<ul style="list-style-type: none"> <li>- 네트워크 메타분석 수행 시 일관성에 대한 보고</li> <li>- 공통대조군 결과값의 연구 간 동질성을 정량적으로 평가</li> <li>- 메타회귀분석 사용 및 IPD 를 이용한 보정 제시</li> </ul>
한국 심평원 2014	<ul style="list-style-type: none"> <li>- 네트워크 메타분석관련 구체적 언급은 없으나 간접비교 수행 시 전이성 가정을 소명하도록 제시</li> </ul>



### 2.3.2. 일관성 확보를 위한 교란요인 보정

인슐린 치료를 받는 대부분의 제2형 당뇨병 환자들의 경우 목표 혈당조절을 위해 인슐린 요법을 강화하게 된다. 인슐린 강화요법은 저혈당의 위험이 증가하고 체중이 증가하는 부작용이 있기 때문에, 이를 대신할 수 있는 치료법으로 인슐린 요법에 다른 경구 당뇨병 약제 및 주사제를 추가하는 병용요법이 고려될 수 있다. 그러나 인슐린 병용요법으로 사용되는 다양한 당뇨병치료제 중 어느 것이 비교 우위에 있는지를 알아보기 위한 임상시험 결과는 전무한 상황이었고, 또한 체계적으로 검토되어 전반적으로 비교된 바도 없었다. 따라서 체계적 문헌 고찰을 기반으로 네트워크 메타분석을 수행하여, 제 2형 당뇨병 환자들 중 인슐린으로 혈당 조절이 안 되는 경우 인슐린 병합요법으로 추가되는 당뇨병 치료제 Dipeptidyl peptidase-4 inhibitor(DPP4i), Glucagon-like peptide-1 receptor agonists(GLP-1RA), Sodium glucose cotransporter-2 inhibitor(SGLT2i), Thiazolidinediones(TZD)의 유효성 및 안전성을 비교 평가하였다[14]. 본 장에서는 이 연구의 1평가변수인 당화혈색소 (glycated hemoglobin, HbA1c)의 치료 전 대비 후의 변화  $\Delta\text{HbA1c}(\%)$ 에 대한 예시를 다뤘다.

체계적 문헌 고찰을 통해 문헌 50편이 선정되었다. 관심 약제들 간 직접비교(head-to-head) RCT는 존재하지 않았고 각 약제들이 공통대조군, 인슐린 또는 인슐린+placebo (이하, 인슐린 혹은 인슐린+placebo 군을 Control로 명명하였다)과 비교한 RCT만 존재하였다. 따라서, 다음 Figure 2에 제시된 바와 같이 개방형 네트워크를 형성하였다. Control군과 DPP4i군을 비교한 RCT의 개수는 15개, Control군과 GLP-1RA군의 RCT는 13개, Control군과 SGLT2i군의 RCT는 9개, Control군과 TZD군의 RCT는 13개였다. 각 연구 특성은 부록 A.1에 제시하였다.

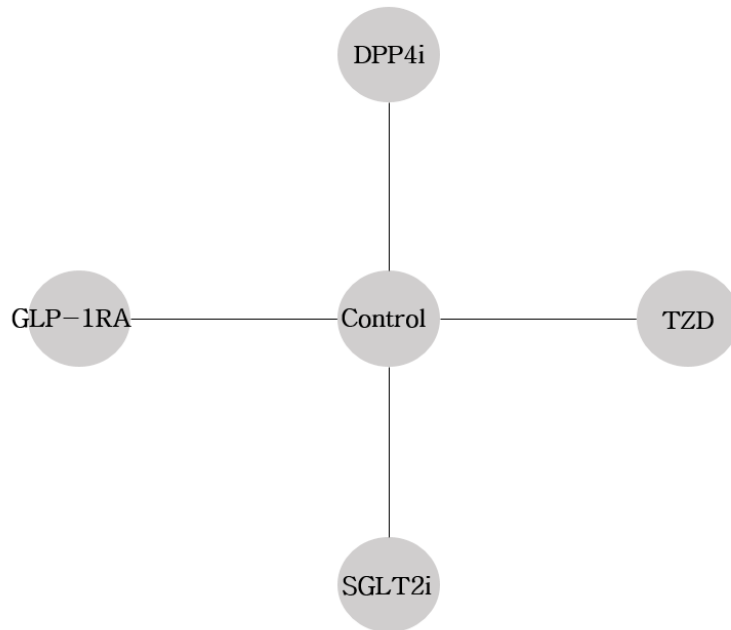


Figure 2. Evidence network for diabetes example

공통대조군(Control군) 결과값에 대한 연구 간 이질성이 상당히 컸다(부록 A.2). DPP4i연구에서 Control군의 결과값 간  $I^2$ 는 94.5%(p-value < 0.001)였고, GLP-1RA연구에서는 97.9%(p-value<0.001), SGLT2i연구에서 94.7%(p-value<0.001), TZD연구에서 80.5%(p-value<0.001)였다. 연구 전체에서 Control군의 결과값 간  $I^2$ 는 95.8%(p-value<0.001)였다. 또한, 다른 연구들에 비해 GLP-1RA연구들의 Control군 효과크기가 큰 경향이 있었다.

Control 군의 결과값은 Control군에 속한 환자들에게 적용되는 인슐린 적정(titration) 방법 등과 같이 명확히 정량화되지 않는 치료의 적극성의 정도와 관련될 수 있기 때문에 효과 크기 조정자(effect modifier)로써 활용될 수 있다. 따라서, Control군의 결과값과 치료효과 크기의 관계를 보기 위해 메타회귀분석을 수행하였고, 이들은 통계적으로 유의한 선형관계가 있었다(p-value=0.02)(Figure 3).

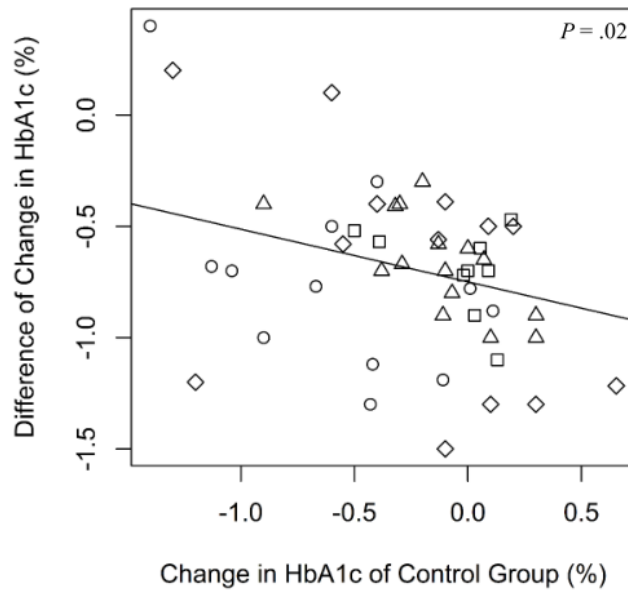


Figure 3. Meta-regression between the change in HbA1c in the control group and the treatment difference in the change of HbA1c.

The triangles, circles, squares, and diamonds correspond to DPP4i studies, GLP-1RA studies, SGLT2i studies and TZD studies, respectively. Solid lines indicate the relationship between the change in the HbA1c in the control group and the treatment difference in the changes of the HbA1c by meta-regression.

측정할 수 있고 임상적으로 유의미할 것으로 판단되는 요인들은  
통합 치료효과 추정에 영향을 끼칠 수 있다. 따라서 다음 제시된 요  
인들을 잠재적 교란요인으로 고려하였다: Age (year), Men  
proportion (%), Body mass index (BMI) at baseline( $\text{kg}/\text{m}^2$ ),  
HbA1c at baseline (%), Diabetes duration (year), Insulin dose at  
baseline (IU). 각 변수에 대해 단변수 메타회귀분석을 수행하였고,  
회귀계수 추정값의 95% 신뢰구간이 '0' 을 포함하지 않는 변수를  
교란요인으로 선정하였다. 통계적으로 유의한 변수는 Age(year),  
Men proportion(%), BMI at baseline( $\text{kg}/\text{m}^2$ ) 였다(Figure 4). 따  
라서, 최종 모형에는 Control군의 결과값, Age (year), Men  
proportion (%), BMI at baseline( $\text{kg}/\text{m}^2$ )가 보정변수로 포함되었다.

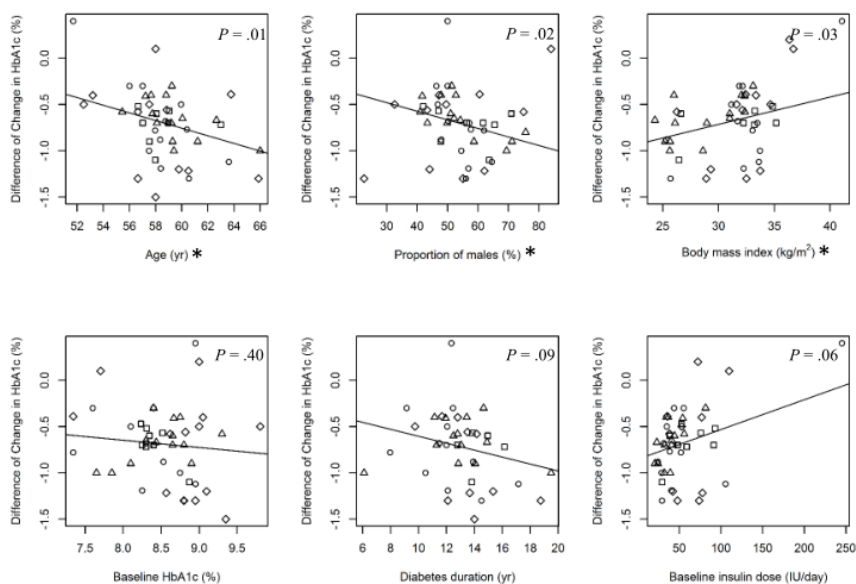


Figure 4. Meta-regression between potential confounders and the treatment difference in the change of HbA1c.

The triangles, circles, squares, and diamonds correspond to DPP4i studies, GLP-1RA studies, SGLT2i studies and TZD studies, respectively. Solid lines indicate the relationship between each potential covariate and the treatment difference in the changes of the HbA1c by meta-regression.

보정 전 혈당 조절에 가장 효과적인 약제는 GLP-1RA였고, 다음은 TZD, SGLT2i, DPP4i 순이었고 이들 간 유의한 차이를 보이는 것은 없었다(부록 A.3).

GLP-1RA연구에서는 다른 연구들에 비해서 공통대조군에서의 결과값이 상대적으로 크게 나타났고(부록 A.2, Figure3) 이를 교란 요인으로 고려하여 보정한 후에 GLP-1RA의 효과크기는 보정 전에 비해 증가함을 확인하였다. 또한, DPP4i연구에 포함된 환자들의 BMI 값은 평균적으로 다른 연구들에 비해 낮았고 이를 보정한 결과를 반영하였을 때는 보정 전에 비해 DPP4i의 효과크기가 감소하였다(Figure 4). 따라서 보정 전과 비교했을 때 약제들의 순위는 동일하였지만 유의한 차이를 보이지 않았던 GLP-1RA와 DPP4i 간의 치료효과 차이가 보정 후 유의한 수준으로 벌어졌다(부록 A.3).



## 2.4. 고찰 및 결론

캐나다, 호주, 영국의 의료기술평가 지침에서는 네트워크 메타분석 결과를 제공할 경우 해당 네트워크의 직접비교 근거와 간접비교 근거의 일관성 탐색 및 보고할 것을 제안하고 있다. 영국 지침의 경우 일관성의 불확실성이 큰 경우 그 원인을 소명하도록 하였다. 캐나다와 호주의 경우, 공통대조군들의 결과값을 활용하여 연구 간 동질성을 평가하고 측정하기 어려운 환자 특성의 대리 지표로서 이를 사용하도록 하였다. 특히 호주 지침의 경우 메타회귀분석 외에도 개별환자데이터를 가용할 수 있는 경우 이를 이용한 보정 방법 사용 등 비일관성의 원인이 될 수 있는 요인들을 정량적으로 파악하고 완화시킬 것을 안내하고 있다.

호주 지침에서 언급된 개별환자데이터를 사용한 보정 방법인 Population-adjusted indirect comparisons[29-31]은 개방형 네트워크 메타분석에 포함된 연구 중 하나 이상의 연구에 대한 개별 환자 데이터를 사용할 수 있는 경우 활용된다. 해당 방법은 연구들 간 환자 특성의 불균형을 개선하기 위해 적용되며, 개별 환자 데이터를 이용하기 때문에 요약 자료를 이용한 메타회귀분석 방법보다 더 정교한 접근법으로 여겨진다. 현재는 세 개의 치료법 이하로 구성된 근거 네트워크(anchored comparisons와 unanchored comparisons)에 한하여 적용되고 있다[34].

국내 경제성평가 지침(심평원, 2011)에서는 간접비교를 수행할

시 포함된 연구 간 전이성 가정을 소명하도록 하였고 네트워크 메타분석에 대해서는 구체적으로 언급하지 않았다. 향후 본 지침은 개편 예정에 있으며 개정판에는 네트워크 메타분석에 대한 제시안이 일부 포함될 예정이다.

서로 간 상호 비교된 바 없는 제2형 당뇨병 환자에게 인슐린 병합요법으로 투약하는 치료제들, DPP4i, GLP-1RA, SGLT2i, TZD 간 비교평가를 위해 교란요인 보정을 적용한 개방형 네트워크 메타분석을 수행하였다. 호주 지침(PBAC 2016)에서 간접비교 수행 시 공통대조군의 결과 값에 대한 동질성을 정량적으로 평가하도록 하여 본 분석에서도 이를 준수하여 공통대조군 간 이질성을  $I^2$ 로 평가하였다. 또한, 캐나다 지침(CADTH 2019)에서는 공통대조군의 결과값을 측정할 수 없는 환자 특성의 대리 지표로 활용하도록 하였고, 본 분석에서도 공통대조군인 Control군의 결과값이 명확히 정량화되지 않는 치료의 적극성의 정도를 반영하고 있기 때문에 이 값을 교란요인으로 고려하였다. 본 분석에서 이질성이 나타난 RCT들로 구성된 개방형 네트워크에 대하여 일관성 가정하에 근거를 통합하였을 때 각 치료제들 간에 불분명하게 나타났던 치료효과 차이는 연구들의 기저 위험 수준을 동일하게 보정함으로써 명확히 구분될 수 있었음을 확인 하였고, 해당 분석 과정을 통해 일관성에 대한 검증이 불가능한 개방형 네트워크에서 전이성 가정을 최대한 만족하도록 하여 잠재적인 일관성 위배를 최소화하는 접근의 중요성을 제시하였다.

## 제3장. 개방형 네트워크 메타분석에 대한

### 체계적 문헌 고찰

#### 3.1. 서론

개방형 네트워크 메타분석에서는 동질성 가정과 전이성 가정을 평가 할 수 있다. 동질성과 전이성 가정을 만족하는 개방형 네트워크의 경우 분석에 포함된 연구들의 공통대조군 결과값 간 차이에서 통계적 유의성은 나타나지 않을 것이다. 가정을 만족하지 않는 경우에는 치료효과에 영향을 미치는 잠재적 교란요인을 평가하고 통계적으로 유의한 요인을 보정하는 네트워크 메타분석을 수행한다.

본 장에서는 네트워크 메타분석 접근을 사용하여 출판된 기존 연구 문헌들 중 개방형 네트워크 메타분석의 빈도 및 해당 연구들의 일관성에 대한 접근 방법과 결론의 한계점 등에 대한 경험적 근거 파악을 위해 수행한 체계적 문헌 고찰을 수행하였다.

## 3.2. 방법

### 3.2.1. 문헌 검색 및 선정

2018년 1월부터 2019년 3월까지 출판된 체계적 문헌 고찰 기반 개방형 네트워크 메타분석 논문을 선정하였다. 선정기준은 1차 평가 변수를 기준으로 중재(intervention) 3개 이상을 RCT로만 비교한 개방형 네트워크 메타분석 논문들을 선정하였다. 문헌 검색은 PubMed를 이용하여 수행하였다.

사용한 검색어는 “(Network [tiab] OR mixed treatment [tiab] OR multiple treatment [tiab] OR mixed comparison [tiab] OR indirect comparison [tiab] OR simultaneous comparison [tiab] ) AND meta-analysis [tiab] AND 2018:2019/03 [dp]”이며, 영문으로 출판된 논문으로 제한하였다.

### 3.2.2. 자료 추출 및 요약

각 문헌에서 1차 평가 변수를 기준으로 다음 자료를 추출하였다 (Table 2). 네트워크에 포함된 중재의 개수와 종류, 평가변수의 종류를 추출하였다. 또한, 네트워크에서 동질성 가정과 전이성 가정 검정 평가를 위해 공통치료군 결과값의 동질성을 검토하였는지 그리고 치료효과에 잠재적 교란요인의 영향을 평가하였는지 여부와 분석방법에 대한 정보를 추출하였다.

이 항목에 따라 연구의 개수를 요약하고 전체 네트워크 중 각 항목에 해당하는 연구 수의 비율을 제시하였다.

Table 2. 자료 항목

구분	변수
개방형 네트워크의 특성	<ul style="list-style-type: none"> <li>· 네트워크에 포함된 중재 개수</li> <li>· 중재의 종류 (drug, surgical, educational, nutritional 등)</li> <li>· 평가변수의 종류 (objective, subjective)</li> </ul>
가정평가	<ul style="list-style-type: none"> <li>· 공통대조군의 효과 크기 간 동질성 평가               <ul style="list-style-type: none"> <li>- 정성적 평가</li> <li>- <math>I^2</math>, Q 통계량을 이용한 정량적 평가</li> </ul> </li> <li>· 치료효과에 미치는 잠재적 교란요인의 영향 평가               <ul style="list-style-type: none"> <li>- 정성적 평가</li> <li>- 네트워크 메타회귀분석, 소그룹분석을 이용한 정량적 평가</li> </ul> </li> </ul>
분석방법	<ul style="list-style-type: none"> <li>· 일관성 가정하에 수행된 기본 네트워크 메타분석</li> <li>· 교란요인을 메타회귀를 통해 보정한 분석</li> <li>· 교란요인에 따른 소그룹별 네트워크 메타분석</li> </ul>

### 3.3. 결과

총 1353 편의 논문이 검색되었다. 문헌 선정 과정은 아래 Figure 5에 제시하였다.

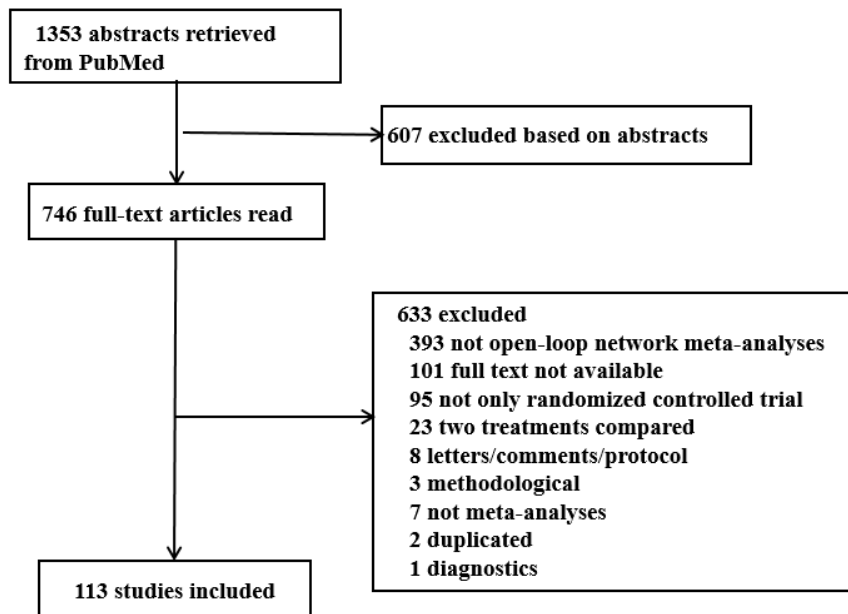


Figure 5. 문헌 선정 과정 흐름도

제목과 초록의 검토를 통해 607 편이 제외되었고, 논문의 전문에 대한 검토를 통해 633편이 제외되었다. 이들 중 393편은 근거 네트워크에 닫힌 루프를 최소 한 개 이상 포함하였기 때문에 제외되었다. 최종적으로, 논문 총 113 편이 선정되었다. 검색 기간 동안 출판된 506개의 네트워크 메타분석 연구 중 개방형 네트워크 메타

분석 연구의 비율이 대략 22.3%임을 확인 할 수 있었다.

각 연구에 포함된 네트워크의 특성은 다음과 같았다(Table 3). 전체 네트워크 중 중재 수가 4개 이하인 네트워크가 40개(35.4%)였고, 5-10개인 네트워크는 62개(59.4%)였고, 그 이상인 경우가 11개(9.7%)였다. 대부분의 네트워크에서 약제 간 비교를 하였으며(94개(83.2%)), 이외에는 수술 간 비교(8개(7.1%)), 영양 간 비교(3개(2.7%)), 교육 간 비교(1(0.9%)), 의료장치 간 비교(1(0.9%)) 혹은 이들의 조합 간 비교(6(5.3%))를 하였다. 또한, 대부분의 네트워크에서 객관적 수치를 기반으로 하는 결과값을 사용하였다(104개(92.0%)).



Table 3. 출판된 개방형 네트워크 특성

특성	No (%) (n=113)
<b>네트워크에 포함된 중재 수*</b>	
-4	40 (35.4%)
5-10	62 (54.9%)
11-	11 (9.7%)
<b>중재 종류</b>	
Drug	94 (83.2%)
Surgical	8 (7.1%)
Nutritional	3 (2.7%)
Educational	1 (0.9%)
Medical device	1 (0.9%)
위 항목의 조합	6 (5.3%)
<b>결과값 종류</b>	
Objective	104 (92.0%)
Subjective	9 (8.0%)

\*공통대조군 포함

개방형 네트워크 메타분석 연구 113 개 중 5개에서 공통대조군의 효과 크기 간 동질성을 평가하였고, 80개(위 5개 연구 포함)의 연구에서 잠재적 교란요인에 대한 평가를 수행하였다(Table 4). 관련 보고를 하지 않은 연구는 33개(29.2%)였다. 공통대조군 효과 크기의 동질성 평가를 정성적으로 수행한 연구는 4개였고 정량적으로 수행한 연구는 1개였다. 또한, 치료효과에 대한 잠재적 교란요인에 대하여 정성적으로 평가한 연구는 66개였고, 정량적으로 수행한 연구는 14개였으며, 사용한 정량적 분석 방법으로는 네트워크 메타회귀분석을 이용한 연구 12개, 소그룹 분석을 이용한 연구가 2개였다.

Table 4. 가정 평가 수행 현황

구분	N
공통대조군 효과 크기 동질성 평가	5
정성적 평가	4
정량적 평가	1
치료효과에 대한 잠재적 교란요인에 대한 평가	80*
정성적 평가	66
정량적 평가	14
Not reported	33
*공통대조군 효과 크기 동질성 평가 문헌 포함	

다음은 분석 방법을 요약하였다(Table 5). 기본 네트워크 메타 분석을 수행한 네트워크는 103개(91.2%)였고, 교란요인을 메타회귀를 통해 보정한 분석을 수행한 네트워크는 6개(5.3%), 교란요인에 따라 구분한 소그룹별 네트워크 메타분석을 수행한 네트워크는 4개(3.5%)였다.

기본 네트워크 메타분석을 수행한 103개의 연구 중 64개에서 각 연구의 저자들이 네트워크의 동질성 및 전이성 가정이 만족됨을 언급하였다. 반면, 6개의 연구에서는 가정을 만족하지 않는다고 언급하였다. 이들 중 3개는 데이터가 충분하지 않았기 때문에 네트워크 메타회귀분석을 수행할 수 없다고 보고하였고 나머지 3개는 보정할 수 없었던 이유를 소명하지 않았다. 그 외 33개 연구에서는 가정 평가를 수행하지 않고 기본 네트워크 메타분석을 수행하였다.

Table 5. 분석 방법 적용 현황

구분	N(%)
일관성 가정 기반의 기본 네트워크 메타분석	103 (91.2%)
동질성 및 전이성 가정을 만족함	64
동질성 및 전이성 가정을 만족하지 않음	6
가정 평가를 수행하지 않음	33
교란요인을 메타회귀를 통해 보정한 분석	6 (5.3%)
교란요인의 소그룹별 네트워크 메타분석	4 (3.5%)
전체	113 (100%)

### 3.4. 고찰 및 결론

PubMed를 통해 확인된 약 1년간 출판된 506개 네트워크 메타분석 연구 중 개방형 네트워크 메타분석 연구는 113개(22.3%)이었고 대부분의 개방형 네트워크 메타분석 연구에서 객관적 수치를 기반으로 하는 결과값을 이용하여 약제 간 비교를 수행하였다.

공통대조군의 결과값은 네트워크에 포함된 연구에서 측정하지 못하였거나 명확하게 드러내기 어려운 요인들을 내포하고 있기 때문에 분석에서 유익하게 활용될 수 있는 자료이다. Cooper et al. 2015[25]에서도 네트워크에 공통대조군의 효과 크기는 측정할 수 없는 환자 특성의 대리 지표로서 이질성 또는 비일관성의 잠재적 원인으로 고려할 수 있다고 제시하고 있다[25]. 또한, 2.3.2에 제시된 비교효과연구에서도 공통대조군 효과 크기가 치료효과 크기 조정자로서 활용되었다. 그러나 본 체계적 문헌 고찰을 통해 파악한 결과, 공통대조군 결과값의 연구 간 동질성을 평가한 연구는 단 5개였다.

전체 113개의 개방형 네트워크 메타분석 연구 중 103개의 연구에서는 일관성 모형을 이용한 기본 네트워크 메타분석을 수행하였다. 이들 중 일관성 가정과 전이성 가정을 평가하고 가정 만족을 확인함을 명시한 연구는 64개였다. 일반적으로 이들은 적절하게 분석된 것으로 고려된다. 그러나, 대부분의 연구에서 가정 평가를 정성적으로 수행하였기 때문에 평가 결과의 객관성에 의문이 제기될

수 있다. 네트워크에 포함된 연구의 수가 적은 경우 가정 평가를 정성적으로 수행하여도 문제가 되지 않을 수 있으나, 네트워크에 포함된 연구가 많은 경우에는 한눈에 관련 정보를 파악하기 어렵기 때문에 정성적인 평가 결과의 신뢰성 확보에 한계가 있을 것이다. 2.3.2에 제시된 예시에서도 연구 간 동질성을 판단하기 위해 연구 및 환자 특성 요약 표(부록 A.1)를 정성적으로 검토하는 것은 불가능할 것이다.

일관성 가정하에 네트워크 메타분석을 수행한 연구 중 6개의 연구에서는 가정을 만족하지 않음을 명시하였다. 이들은 자료의 수가 부족하여 교란요인 보정을 수행할 수 없었음을 보고하였다. 앞서 설명한 호주 지침(PBAC, 2016) [28]에서도 네트워크 메타회귀분석을 수행하기 위해 공변량 변수당 10개의 RCT를 권장하고 있다.

전체 연구 중 96개 연구는 정성적으로 가정 평가를 수행하여 평가 결과의 객관성이 부족한 경우(57개), 자료 수가 부족하여 교란요인을 보정할 수 없는 경우(6개), 가정 평가를 수행하지 않은 경우(33개)였으나, 모두 일관성 가정하에서 자료 통합을 수행하였다. 이처럼 출판된 대다수의 개방형 네트워크 메타분석들은 확신할 수 없는 일관성 가정하에 간접비교결과에만 기반하여 통합된 결과가 제시되고 있었음을 확인할 수 있었고, 본 정성적 체계적 문헌고찰을 통해 이 들 연구에 실제적으로 직접비교 연구 자료가 포함될 수 있었다면 기존 결론이 유지될 수 있을 것인가에는 심각한 불확실성이 존재하며 이러한 불확실성이 최종적인 결론 및 의료 의사결정에 반

영에 될 수 있어야 함에 대한 문제를 제기하였다.



## 제4장. 개방형 네트워크 메타분석 결과의

### 신뢰성 평가 방법

#### 4.1. 서론

네트워크 메타분석에서 일관성 가정을 만족하지 않는 경우, 그 원인을 소명하기 위하여 잠재적 요인들에 대해 정성적 및 정량적 평가가 필요하다[18, 35, 36]. 잠재적 요인으로는 연구 단위의 환자들의 특성, 비뮌림 위험 요인, 각 연구로부터 추출된 효과 크기에 대한 정확도 등이 있으며, 통계적으로 유의한 요인들을 선정하여 이들을 보정한 네트워크 메타분석을 수행함으로써 자료 통합에 대한 타당성이 확보될 수 있도록 한다[15, 36, 37]. 일관성 가정을 만족하는 경우에는 일관성 모형을 이용한 네트워크 메타분석을 수행한다[10, 38, 39]. 이 분석에서는 통계적으로 허용하는 비일관성의 정도가 연구 간 이질성의 일부로 취급되어 효과 크기의 통합추정치와 그 분산 값에 반영된다.

개방형 네트워크에 포함된 연구의 특성이 충분히 동질하지 않은 경우, 연구 간의 교환성가능성의 결여를 야기할 수 있는 요인을 보정하는 것이 최선의 방법으로 고려된다[14, 40, 41]. 또한, 통계적으로 연구 간에 교환이 가능하다고 판단될 때는 일관성 모형을 이용한 간접비교에 기초하여 네트워크 메타분석을 수행한다[42-44]. 그러나 제3장에 제시된 체계적 문헌 고찰을 통해 대다수의 개방형

네트워크 메타분석들은 확신할 수 없는 일관성 가정하에 간접비교 결과에만 기반하여 통합된 결과가 제시되고 있었음을 확인하였다. 따라서, 본 장에서는 근거 네트워크에 포함된 연구들 간 특성의 정성적 평가만을 통해 일관성이 충족됨을 가정하여 근거 통합이 수행되는 개방형 네트워크 메타분석에서 얻어지는 결과의 신뢰성 평가를 위한 방법론을 제시하였다.

## 4.2. 방법

### 4.2.1. 통계적 방법

Figure 6는 공통 대조군에 해당하는 치료법( $T_0$ )과  $p$  개의 활성 치료법들( $T_1, \dots, T_p$ )로 구성된 개방형 네트워크이다(Figure 6). 이 경우 활성 치료법들 간 비교 수는  $p(p-1)/2$  개이며, 이들에 대한 RCT는 존재하지 않는다. 개발한 분석 방법은 개방형 네트워크에 연결되지 않은  $p(p-1)/2$  개의 비교에 대한 가상의 RCT 데이터를 생성하여 모든 치료법들이 서로 연결된 네트워크를 만들어 네트워크 메타분석을 수행하는 것이다. 이하, 모든 치료법들이 서로 연결된 네트워크를 ‘완전한 네트워크(complete network)’라 명명하였다. 자료산입(算入, imputation)을 위해 생성한 데이터는 연구-단위의 치료효과 크기( $\hat{\theta}_{ijk}$ )와 이 값의 분산( $\widehat{var}(\hat{\theta}_{ijk})$ )이며, 통계적으로 수용 가능한 정도의 비일관성이 고려되어 자료산입 후 생성된 완전한 네트워크의 자료들이 일관성 가정하에 통합될 수 있도록 시뮬레이션 하였다. 본 방법을 ‘자료산입을 통한 보정법(adjustment through imputation)’으로 명명하였다. 개방형 네트워크 메타분석 결과와 자료산입을 통한 보정법을 이용한 민감도 분석 결과의 일치도(agreement)를 비교하여, 개방형 네트워크로부터 얻은 결과의 신뢰성을 평가하였다.

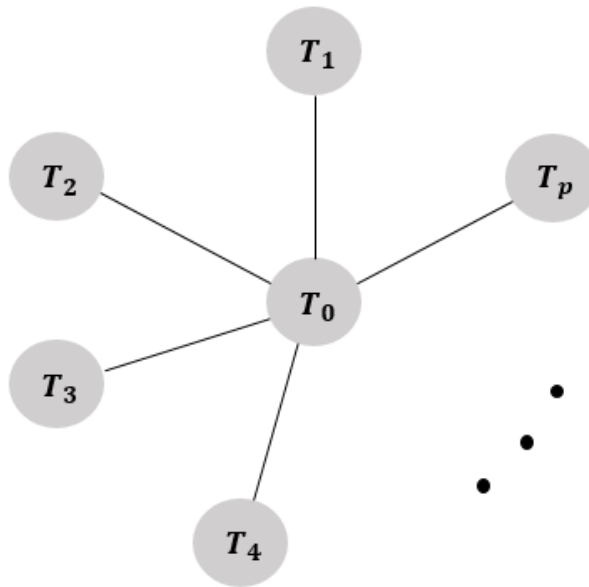


Figure 6. A graphical representation of an open-loop network consisting of one common comparator treatment ( $T_0$ ) and  $p$  active treatments ( $T_1, \dots, T_p$ ).

#### 4.2.1.1. 자료산입 전략

Figure 6에 제시된 네트워크에서  $p$  개의 활성 치료법들 사이의  $p(p-1)/2$ 개 비교에 대한 가상의 데이터는 다음 4가지 조건하에 생성되었다.

- I: 활성 치료법들 간 각 비교에 대해, 개방형 네트워크 메타분석에서 추정된 효과 크기가 양수(positive)이면, 자료산입을 위해 생성된 데이터들의 통합추정치가 개방형 네트워크 메타분석에서 추정된 값보다 작다. (개방형 네트워크 메타분석에서 추정된 효과 크기가 음수(negative)이면, 자료산입을 위해 생성한 데이터들의 통합추정치가 개방형 네트워크 메타분석에서 추정된 값보다 크다.)

보수적인 관점에서 민감도 분석을 실행하기 위하여 가상의 직접비교 추정치가 개방형 네트워크 메타분석에서 관찰된 간접비교 추정치보다 작을 (또는 클) 것이라는 가정을 정립하였다.

- II: 활성 치료법들 간 각 비교에 대해, 자료산입을 위해 생성된 데이터들의 통합추정치의 정확도(extent of precision)는 개방형 네트워크 메타분석에서 추정된 효과 크기의 정확도와 같다.
- III: 활성 치료법들 간 각 비교에 대해, 자료산입을 위해 생성된 데이터를 구성하는 효과 크기의 연구 간 이질성의 정도는

개방형 네트워크의 전체적인 이질성의 정도와 같다.

- IV: 자료산입을 통해 생성한 완전한 네트워크를 활용하여 일관성 모형에 적합하여 얻은 residual deviance 값이 비일관성 모형에 적합하여 얻은 residual deviance 값보다 작다 ([2.2.3. 일관성 가정 검정 방법] 참고). 단, 본 연구에서는 두 모형으로부터 얻은 residual deviance 간 의미 있는 차이를 결정하는 기준치(cut-off value)를 고려하지 않았다.

#### 4.2.1.2. 자료산입을 통한 보정법

A와 B의 RCT들과 A와 C의 RCT를 포함하는 가장 간단한 개방형 네트워크를 사용하여 개발한 방법을 설명하였다.

A와 B의 RCT 데이터는  $\hat{\theta}_{iAB}$ 와  $\widehat{var}(\hat{\theta}_{iAB})$ ,  $i = 1, \dots, N$  ( $N$ 은 A와 B의 RCT의 개수)이고, A와 C의 RCT 데이터는  $\hat{\theta}_{jAC}$ 와  $\widehat{var}(\hat{\theta}_{jAC})$ ,  $j = 1, \dots, M$  ( $M$ 은 A와 C의 RCT의 개수)이다.

위 데이터를 사용한 네트워크 메타분석으로부터 기본 모수에 대한 추정치  $\hat{d}_{AB}$ 과  $\hat{d}_{AC}$ 을 얻고, 네트워크 내 모든 비교들에 대해 공통의 연구 간 변동 추정치  $\hat{\tau}^2$ 를 얻었다. B에 대한 C의 상대적 효과 크기 추정치는  $\hat{d}_{AC} - \hat{d}_{AB}$ 이고 이 값의 분산은  $\widehat{Var}(d_{AC} - d_{AB})$ 이다.

가상의 B와 C 비교 RCT 데이터는  $\hat{\theta}_{kBC}^*$ ,  $\widehat{var}(\hat{\theta}_{kBC}^*)$ ,  $k = 1, \dots, l$  ( $l$ 은 B와 C의 RCT의 개수)로 표기하였고, [4.2.1.1 자료산입 전략]을 기반으로 생성하였다.

$\hat{\theta}_{1BC}^*, \dots, \hat{\theta}_{lBC}^*$ 은 다음 분포(§)로부터 생성하였다:

$$\hat{\theta}_{1BC}^* \sim N(\theta_{1BC}^*, \widehat{var}(\hat{\theta}_{1BC}^*))$$

$$\hat{\theta}_{2BC}^* \sim N(\theta_{2BC}^*, \widehat{var}(\hat{\theta}_{2BC}^*))$$

⋮

$$\hat{\theta}_{lBC}^* \sim N(\theta_{lBC}^*, \widehat{var}(\hat{\theta}_{lBC}^*)).$$

$\theta_{1BC}^*, \dots, \theta_{lBC}^*$ 은 자료산입 모수(imputation parameter)이며, 이들은

평균이  $\hat{d}_{AC} - \hat{d}_{AB} + \bar{\omega}_{BC}$  이고 분산이  $\hat{\tau}^2$  인 정규분포  $N(\hat{d}_{AC} - \hat{d}_{AB} + \bar{\omega}_{BC}, \hat{\tau}^2)$ 로부터 무작위로(random) 생성되었다.  $\bar{\omega}_{BC}$ 은 B와 C의 비교에 대한 가상의 직접비교 근거( $\hat{\theta}_{kBC}^*$ )와 간접비교 근거( $\hat{d}_{AC} - \hat{d}_{AB}$ ) 간 비일관성의 정도를 나타내는 값이며, 이 값은 임의의 상수로 정의된다. 조건 I 을 기반으로, 만일  $\hat{d}_{AC} - \hat{d}_{AB} < 0$ 이면,  $\bar{\omega}_{BC}$ 는 양수로 설정되며,  $\hat{d}_{AC} - \hat{d}_{AB} > 0$ 이면  $\bar{\omega}_{BC}$ 는 음수로 설정된다.

$\widehat{var}(\hat{\theta}_{kBC}^*)$ ,  $k = 1, \dots, l$ 는 조건II와 조건III이 만족될 수 있도록 유도된 다음 공식  $\widehat{var}(\hat{\theta}_{1BC}^*) = \dots = \widehat{var}(\hat{\theta}_{lBC}^*) = l \cdot \widehat{var}(\hat{d}_{AC} - \hat{d}_{AB}) - \hat{\tau}^2$ 을 계산하여 생성하였다(이 공식의 유도 과정은 부록 B.1에서 찾아볼 수 있다). 과정을 간결성을 위해 이들을 모두 동일한 값으로 가정하였다. 단,  $l$ 는 제한조건 ' $l \cdot \widehat{var}(d_{AC} - d_{AB})$ 가  $\hat{\tau}^2$ 보다 크다'를 만족하도록 하는 임의로 숫자이다.

분석에서 자료산입 데이터의 불확실성을 반영하기 위해 다중대체법(multiple imputation method)의 개념[45, 46]이 사용되었다. 다중대체법 방법을 적용하여 정의된 자료산입 데이터의 분포( $\S$ )에서  $\theta_{1BC}^*, \dots, \theta_{lBC}^*$ 를 500번 반복 추출하여 500개의 완전한 네트워크를 생성하고, 각 네트워크별로 일관성 모형을 사용하여 네트워크 메타 분석을 수행하였다. 분석에서 얻은 모수들( $d_{AC}$ ,  $d_{AB}$ ,  $d_{AC} - d_{AB}$ )의 추정치와 치료군별 최우선 순위로 선택되어질 확률의 추정치를 Rubin' s rule에 의해 통합하였다[45].

위 과정은  $|\bar{\omega}_{BC}|$ 값을 0에서부터 점차적으로 증가시키며 조건IV를



만족시키는 범위 내에서 반복적으로 수행되었으며, 이를 통해 조건 IV를 만족하도록 하는  $|\bar{\omega}_{BC}|$ 값의 범위가 도출 되었다. 이 범위는 통계적으로 허용할 수 있는 잠재적 비일관성 크기의 범위를 의미한다. 이 범위 내에서 수행된 민감도 분석에서 도출된 치료법들의 순위가 원래의 개방형 네트워크로부터 도출된 순위와 같도록 하는  $|\bar{\omega}_{BC}|$ 값의 비율을 퍼센트 값으로 제시하였다. 이 값은 잠재적 비일관성에 따른 개방형 네트워크 메타분석 결과의 강건성(robustness)의 정도에 대한 지시값(indicator)으로 사용될 수 있다.

위 방법이 일반적인 개방형 네트워크에서도 적용될 수 있도록 확장하였다. 활성 치료군 개수가  $p$ 개인 개방형 네트워크의  $p(p-1)/2$ 개의  $\bar{\omega}_{jk}$ ,  $j = T_1, \dots, T_{p-1}$  and  $k = T_2, \dots, T_p$  ( $j \neq k$ )가 각각 조건 I가 충족되는 방향으로 동시에 동일한 크기로 변화된다. 이해를 돕기 위해 아래 예시에서는  $p = 3$ 인 개방형 네트워크에 대한 민감도 분석 방법 적용을 설명하였다.

위 방법은 통계 소프트웨어 R에서 패키지 R2WinBUGS를 사용하여 구현되었다[47, 48].

## 4.2.2. 데이터 적용

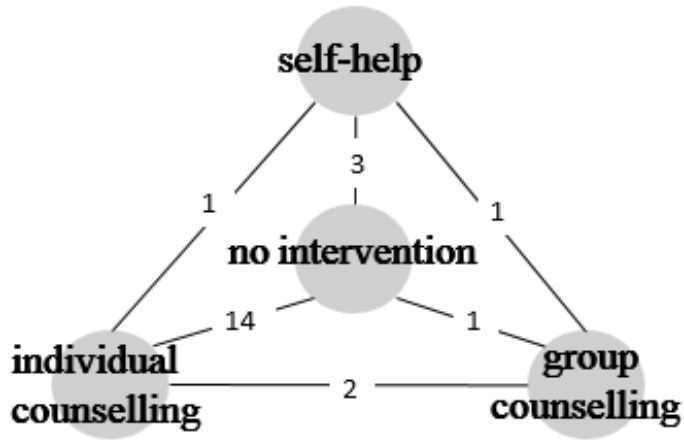
위 방법을 실제 데이터에 적용하여 본 방법의 타당성을 확인하였고, 치료효과 크기의 연구-내 혹은 연구 간 변동성 크기에 따라 생성한 가상의 개방형 네트워크에 적용하고 결과를 비교하여 적용성을 확인하였다.

### 4.2.2.1. Smoking cessation data set

금연 치료 상담 프로그램 No Intervention (NI), Self-Help (SH), Individual Counselling (IC), Group Counselling (GC) 간 효과를 비교한 연구이다[49]. 결과값은 6~12개월 이내 금연 성공 여부이며, log-odds ratio(logOR)가 상대 효과 크기의 측도로 사용되었다. 모든 중재 간 RCT가 존재하였고, RCT의 개수는 세군 비교 연구 두 개를 포함하여 총 24개였다.

본 연구에서는 과정의 편이를 위해 두 군 비교 RCT 22개만 사용하였다. NI와 SH의 RCT 3개, NI와 IC의 RCT 14개, NI와 GC, SH와 IC, SH와 GC의 RCT는 각 1개이다(Figure 7(a)). 이 데이터를 사용하여 일관성 모형을 이용한 네트워크 메타분석(모형의 참조군은 NI임)을 수행하였다. 분석에서 기본 모수  $d_{NI,SH}$ ,  $d_{NI,IC}$ ,  $d_{NI,GC}$ 와 치료군별 최우선 순위로 선택될 확률을 추정하였다.

(a)



(b)

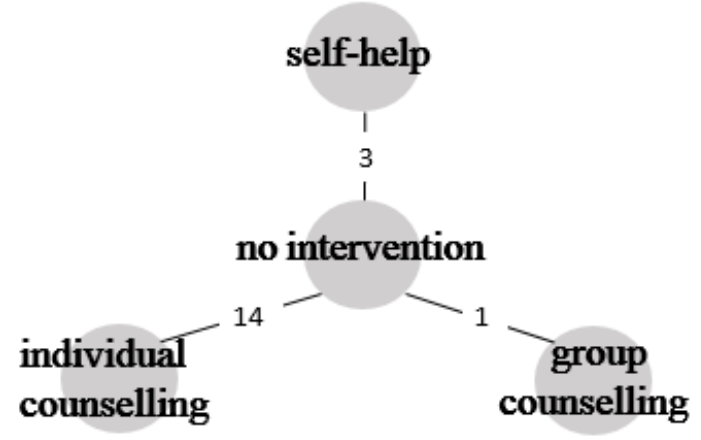


Figure 7. Example networks for smoking cessation data set

(a) A graphical representation of the evidence network for four smoking cessation counseling programs. (b) A graphical representation of the derived open-loop network by eliminating four trials corresponding to direct comparisons among self-help, individual counseling, and group counseling.

다음 Figure 7(b)는 SH, IC, GC 간 RCT들을 제거하여 만든 개방형 네트워크이다. 이 개방형 네트워크를 구성하는 데이터를 사용하여 일관성 모형을 이용한 네트워크 메타분석과 자료산입을 통한 보정법을 이용한 민감도 분석을 수행하였다. 각 분석에서 기본 모수  $d_{NI,SH}$ ,  $d_{NI,IC}$ ,  $d_{NI,GC}$ 와 치료군별 최우선 순위로 선택될 확률을 추정하였다.

자료산입을 통한 보정법을 이용한 민감도 분석에서 비일관성 크기의 절대값  $|\omega_{jk}|$  값의 간격은 0.1로 설정하였고,  $|\omega_{jk}|$  ( $j = SH, IC$ , and  $k = IC, GC$ ,  $j \neq k$ )에 따른 일관성 모형과 비일관 모형의 residual deviance 값들의 변화를 산점도 그래프로 나타내고, 두 모형의 residual deviance 값이 교차되는 점을 표시하여 통계적으로 허용 가능한 비일관성 크기 범위를 얻었다. 또한, 이 범위 내에서 도출된 기본 모수의 추정치와 추정치의 95% 신뢰구간(CI, Confidence interval)을 구간 그래프(interval plot)로 제시하고, 금연 프로그램별 최우선 순위로 선택될 확률의 추정치를 추세 그래프(trend plot)로 나타냈다. 마지막으로, 이 범위 내에서 도출된 금연 프로그램의 순위가 원래의 개방형 네트워크로부터 도출한 순위와 같도록 하는  $|\omega_{jk}|$  값의 비율을 퍼센트 값으로 제시하였다. 관련 코드는 부록 B.2에 수록하였다.

#### 4.2.2.2. Simulated data sets

치료효과 크기의 연구-내 및 연구 간 변동성 수준(즉, 각 연구의 치료효과 크기 값의 표준 오차와 치료효과 크기의 연구 간 이질성의 정도를 말함)에 따라 A와 B의 RCT와 A와 C의 RCT로 구성된 가상의 개방형 네트워크에 대한 데이터를 생성하였다(Table 6). 각 데이터 세트를 A와 B의 RCT 5개와 A와 C의 RCT 5개로 구성하였다. 각 비교의 치료효과 크기 값들은  $I^2$ 통계량을 기준으로 임의로 선택하였으며,  $I^2$  통계량은 0%(no heterogeneity)과 40%(moderate heterogeneity)로 고려되었다[50]. 세 군간 효과 크기를 구별하기 위하여, C, B, A 순으로 A에 대한 B의 통합 치료효과 크기가 0.5가 되도록, A에 대한 C의 통합 치료효과크기가 1이 되도록 설정하였다. 각 데이터 세트별로 치료효과 크기 값의 표준 오차 수준을 고정된 값 2 또는  $2\sqrt{2}$ 로 동일하게 설정하였다.

각 데이터 세트에 개발한 자료산입을 통한 보정법을 적용하였다.  $|\omega_{jk}|$  값의 간격은 0.1로 설정하였고,  $|\omega_{AB}|$ 에 따라 일관성 모형과 비일관 모형의 residual deviance 값들의 변화를 산점도 그래프로 나타내고, 두 모형의 residual deviance 값이 교차되는 점을 표시하여 통계적으로 허용 하는  $|\omega_{AB}|$ 의 범위를 얻었다. 이 범위 내에서 치료군별 최우선 순위로 선택될 확률의 추정치를 추세 그래프로 제시하였다. 최종적으로, 위 범위 내에서 수행한 민감도 분석에서 원래의 순위(C, B, A 순)와 동일한 순위를 도출하도록 하는  $|\omega_{AB}|$  값

의 비율을 백분율 값으로 제시하였다. 관련 코드는 부록 B.3에 수록하였다.

Table 6. Simulated data sets

Study	Comparison	(a)		(b)		(c)		(d)	
		Effect size	Standard error	Effect size	Standard error	Effect size	Standard error	Effect size	Standard error
1	A vs B	0.5	2	-2.5	2	0.5	$2\sqrt{2}$	-4	$2\sqrt{2}$
2	A vs B	0.5	2	-1.5	2	0.5	$2\sqrt{2}$	-2	$2\sqrt{2}$
3	A vs B	0.5	2	0.5	2	0.5	$2\sqrt{2}$	0.5	$2\sqrt{2}$
4	A vs B	0.5	2	2.5	2	0.5	$2\sqrt{2}$	3	$2\sqrt{2}$
5	A vs B	0.5	2	3.5	2	0.5	$2\sqrt{2}$	5	$2\sqrt{2}$
6	A vs C	1	2	-2	2	1	$2\sqrt{2}$	-3.5	$2\sqrt{2}$
7	A vs C	1	2	-1	2	1	$2\sqrt{2}$	-1.5	$2\sqrt{2}$
8	A vs C	1	2	1	2	1	$2\sqrt{2}$	1	$2\sqrt{2}$
9	A vs C	1	2	3	2	1	$2\sqrt{2}$	3.5	$2\sqrt{2}$
10	A vs C	1	2	4	2	1	$2\sqrt{2}$	5.5	$2\sqrt{2}$

(a) when  $I^2$  is 0% and the standard error is 2, (b) when  $I^2$  is 40% and the standard error is 2, (c) when  $I^2$  is 0% and the standard error is  $2\sqrt{2}$ , and (d) when  $I^2$  is 40% and the standard error is  $2\sqrt{2}$ .

## 4.3. 결과

### 4.3.1. Smoking cessation data set

원래의 네트워크(Figure 7(a))에서  $\hat{d}_{NI,SH}^c$ ,  $\hat{d}_{NI,IC}^c$ ,  $\hat{d}_{NI,GC}^c$ 은 각각 0.43 (95% CI, -0.38 to 1.25), 0.73 (0.26 to 1.20), 1.38 (0.25 to 2.5)였고, 금연 성공에 가장 효과적인 치료법은 GC이고 그 다음은 IC, SH, NI 순이었다(Table 7). 활성 치료군 간 RCT가 제외된 개방형 네트워크(Figure 7(b))에서는  $\hat{d}_{NI,SH}^s$ ,  $\hat{d}_{NI,IC}^s$ ,  $\hat{d}_{NI,GC}^s$ 가 각각 0.33 (-0.73 to 1.39), 0.72 (0.19 to 1.25), 3.52 (0.12 to 6.93)이었으며, 치료법들의 순위는 원래의 네트워크로부터 얻은 것과 동일하게 도출되었다. 하지만, 개방형 네트워크에서 얻은 NI에 대한 GC의 치료효과 크기 값( $\hat{d}_{NI,GC}^s$ )이 원 네트워크에서 얻은 값( $\hat{d}_{NI,GC}^c$ )보다 더 컸으며, GC가 최우선으로 선택될 확률도 높았다.



Table 7. Results of a network meta-analysis using the complete network and derived open-loop network for the smoking cessation data

	Parameter	Complete network		Open-loop network	
		Estimate (95% CI)	P(best)	Estimate (95% CI)	P(best)
Treatment					
A: NI	-		0.00		0.00
B: SH	$d_{AB}$	0.43 (-0.38, 1.25)	0.04	0.33 (-0.73, 1.39)	0.02
C: IC	$d_{AC}$	0.73 (0.26, 1.20)	0.10	0.72 (0.19, 1.25)	0.04
D: GC	$d_{AD}$	1.38 (0.25, 2.50)	0.86	3.52 (0.12, 6.93)	0.94
Heterogeneity	$\tau$	0.78		0.85	

CI, confidence interval; NI, no intervention; SH, self-help; IC, individual counseling; GC, group counseling

해당 개방형 네트워크의 자료산입을 통한 보정법을 이용한 민감도 분석에서 통계적으로 허용하는  $|\omega_{jk}|$  값의 범위는 0부터 1.05까지였다(Figure 8).

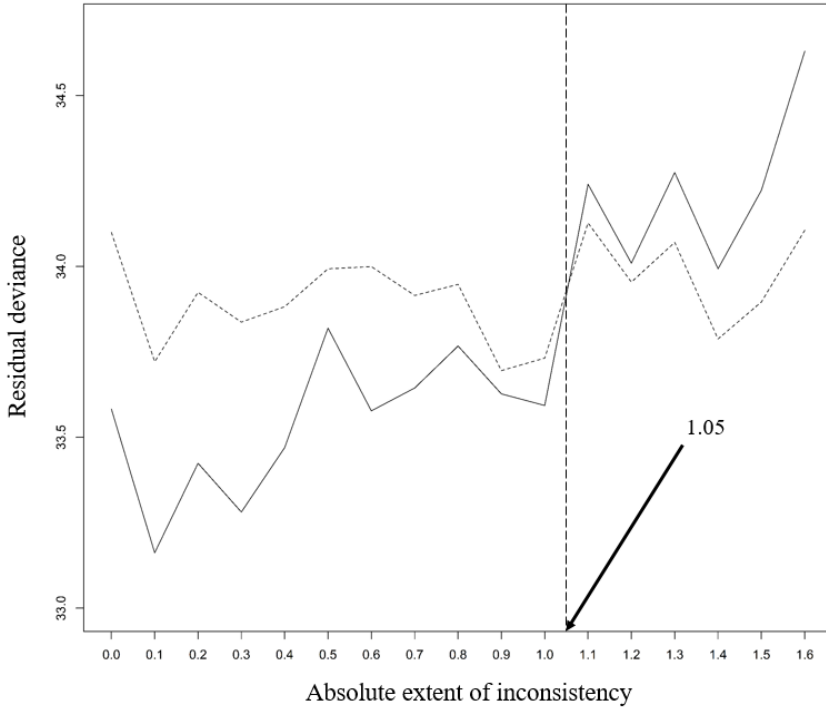


Figure 8. Residual deviances by model type (y-axis) against the absolute extent of inconsistency (x-axis)

The solid line and dashed line indicate the consistency model and inconsistency model, respectively. A vertical line marks the point at which the two lines cross, and the value of that point on the x-axis is shown.

비일관성의 크기의 절대값이 증가하면서  $d_{NI,SH}$ 의 추정치는 증가,  $d_{NI,GC}$ 의 추정치는 감소하여 추정치들 간 차이가 줄어들었다(Figure 9).

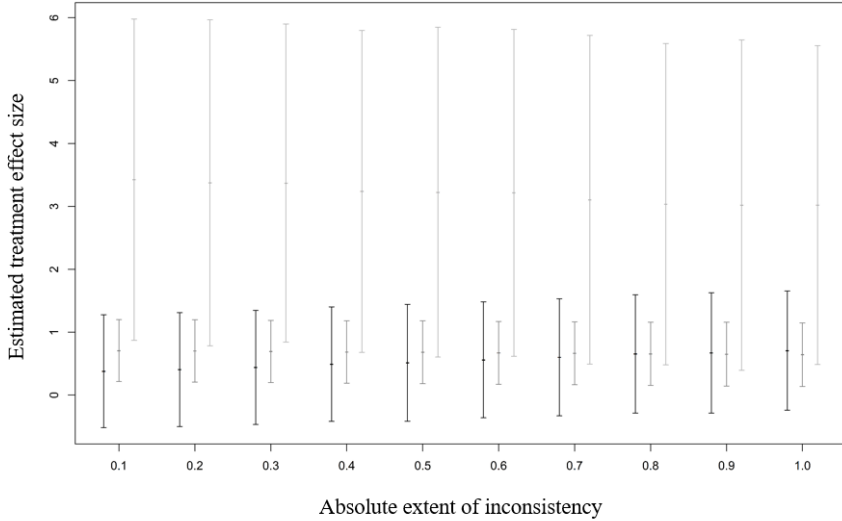


Figure 9. Interval plot of estimates of basic parameters against the extent of inconsistency (x-axis) within the obtained range.

Black, gray, and dim gray lines indicate the estimated treatment effect sizes for self-help, individual counseling, and group counseling compared to no intervention with the accompanying 95% confidence intervals, respectively.

통계적으로 허용하는  $|\omega_{jk}|$  값 범위 내에서 개방형 네트워크에서 과장되었던 GC가 최우선으로 선택될 확률은 자료산입을 통한 보정법을 이용하여 원래의 완전한 네트워크에서 얻은 결과와 비슷한 수준으로 감소하였다(Figure 10 (a)). 또한, 해당 범위내에서 개방형 네트워크 메타분석에서 얻은 순위와 같도록 하는  $|\omega_{jk}|$ 의 비율은 약 67%였다(Figure 10 (b)) (이 비율은  $100 * 0.7 / 1.05$  로 계산된다).

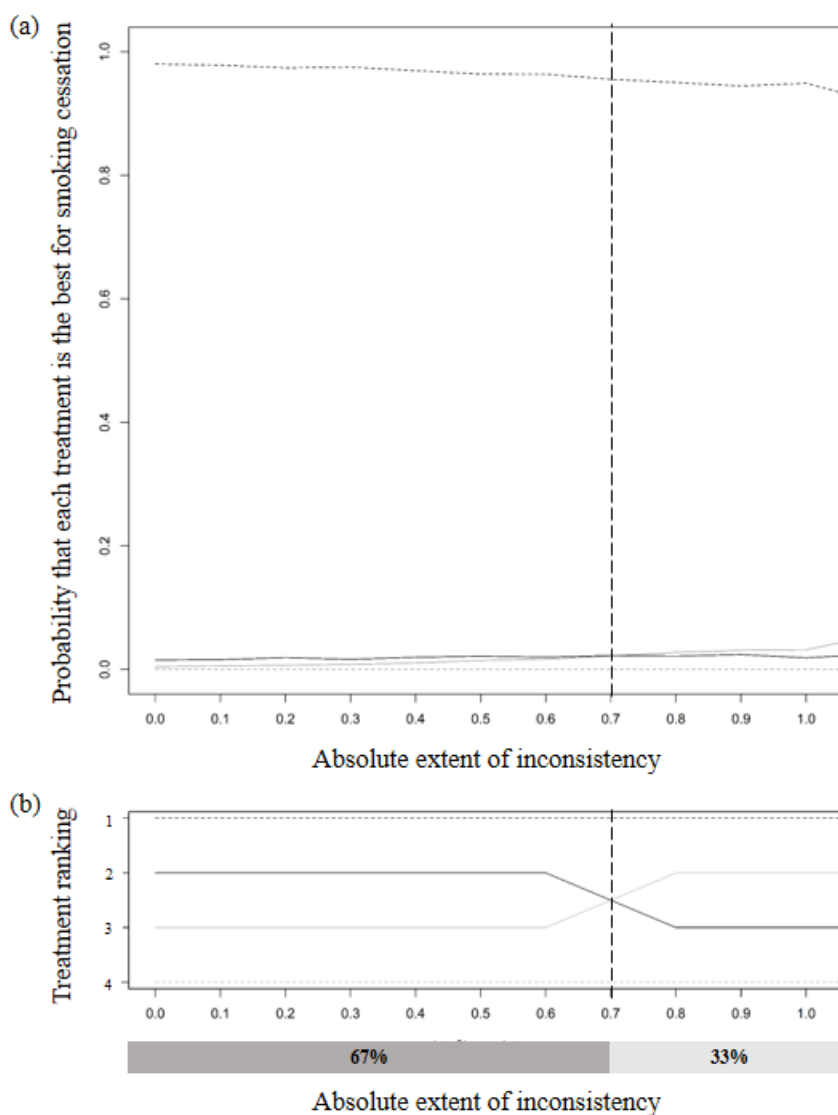


Figure 10. (a) Probability that each treatment is the best for smoking cessation against the extent of inconsistency within the obtained range. (b) Ranking of each treatment for successful smoking cessation against the extent of inconsistency within the obtained range.

The gray dotted, gray solid, black solid, and black dotted lines

indicate the probabilities corresponding to no intervention, self-help, individual counseling, and group counseling, respectively. A vertical line marks the point at which some lines cross, and the percentages in the dark gray and dim gray boxes represent the proportions of  $|\omega_{jk}|$  that resulted in a consistent ranking and an inconsistent ranking of treatments relative to the original ranking, respectively.

### 4.3.2. Simulated data sets

치료효과 크기의 연구-내 변동 수준이 동일할 때 연구 간 변동 수준이 클수록 일관성 모형과 비일관성 모형의 두 residual deviance 곡선이 교차하는  $|\omega_{AB}|$ 의 값이 더 커졌다(Figure 11). 또한, 연구 간 변동 수준이 동일할 때 연구-내 변동이 클수록 두 모형의 residual deviance 곡선이 교차하는  $|\omega_{AB}|$ 의 값이 더 커졌다. 이 양상은 본 연구에서 제안한 자료산입을 통한 보정법이 근거 네트워크의 불확실성의 정도가 클수록 통계적으로 허용하는 비일관성 수준이 더 크다는 것을 반영하고 있음을 보여준다.

또한, 교차되는  $|\omega_{AB}|$  값이 클수록 범위 내에서 설정한 치료법들의 순위(C, B, A 순)와 동일한 결과를 갖는  $|\omega_{AB}|$ 의 비율은 더 작았다(Figure 12). 산출된 비율이 작다는 것은 일관성 모형을 이용할 수 있는 완전한 네트워크에서 얻은 결과가 일부 치료군 간 RCT가 존재하지 않을 때, 즉 만일 활성 치료군 간 RCT가 존재하지 않는 개방형 근거 네트워크였다면, 통합한 결과와 달라질 가능성이 높다는 것을 나타낸다.

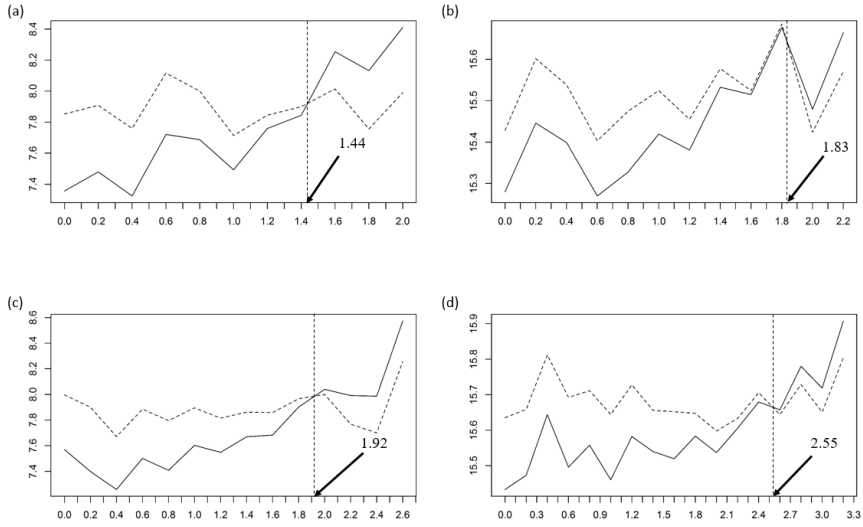


Figure 11. Residual deviances by model type (y-axis) against the absolute extent of inconsistency (x-axis) for each simulated data set.

(a) when  $I^2$  is 0% and the standard error is 2, (b) when  $I^2$  is 40% and the standard error is 2, (c) when  $I^2$  is 0% and the standard error is  $2\sqrt{2}$ , and (d) when  $I^2$  is 40% and the standard error is  $2\sqrt{2}$ . The solid line and dashed line indicate the consistency model and inconsistency model, respectively. A vertical line marks the point at which the two lines cross.



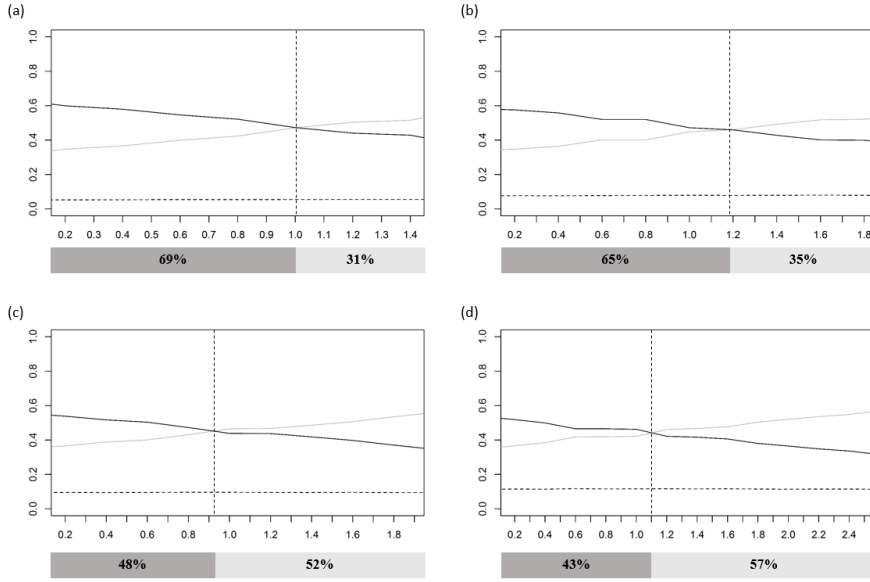


Figure 12. Probability of each group being the best intervention (y-axis) against the extent of inconsistency,  $|\omega_{BC}|$ , (x-axis) within the obtained range for each data set.

(a) when  $I^2$  is 0% and the standard error is 2, (b) when  $I^2$  is 40% and the standard error is 2, (c) when  $I^2$  is 0% and the standard error is  $2\sqrt{2}$ , and (d) when  $I^2$  is 40% and the standard error is  $2\sqrt{2}$ . The black dotted, gray solid, and black solid lines indicate the probability corresponding to groups A, B, and C, respectively. A vertical line marks the point at which some lines cross, and the percentages in the dark gray and dim gray boxes represent the proportions of  $|\omega_{BC}|$  that resulted in a consistent ranking and an inconsistent ranking of treatments relative to the

original ranking, respectively.

#### 4.4. 고찰 및 결론

본 연구에서 개방형 네트워크 메타분석에 대한 민감도 분석을 수행하기 위한 자료산입 방법을 이용한 보정법을 제안하였다. 자료산입 데이터는 네 가지 조건을 기반으로 생성된다. 조건 I 은 보수적인 관점에서 데이터를 생성하기 위해 설정하였다. 치료효과 차이의 참값(true value)은 null에 가깝거나 효과의 방향이 관찰된 값의 반대 방향에 위치하는 것으로 가정하여 관찰된 효과 크기를 편중된 것으로 간주하였다. 이 조건을 설정하여 보수적인 관점에서 민감도 분석을 실행하였다. 또한, 일반적으로 간접적으로 추정한 효과 크기가 직접비교 추정치보다 더 큰 분산을 갖기 때문에[10], 조건 II 에서 자료산입을 위해 생성된 각 직접비교 효과 크기의 분산 값이 간접비교 효과 크기의 분산 값과 같도록 하였다. 조건 III 으로 자료산입 전과 후 네트워크의 전체적인 이질성 수준을 유지하도록 하였다. 따라서 자료산입 방법을 통해 오직 잠재적 비일관성 크기에 따른 결과의 변화를 관찰할 수 있었다. 마지막으로, 자료산입을 통해 생성된 완전한 네트워크를 일관성 모형을 이용하여 통합할 수 있도록 조건 IV 를 설정하였다. 이는 개방형 네트워크 메타분석으로부터 얻은 간접비교 결과와 자료산입을 위해 생성한 직접비교 결과의 차이가 통계적으로 유의하지 않도록 하여 두 결과의 통합이 허용되는 상황을 시뮬레이션하기 위한 조건이다.

본 방법에서는 한 개의 공통대조군과  $p$  개의 활성 치료법들로 구

성된 개방형 네트워크에 대하여, RCT의 부재로 연결되지 않은  $p(p-1)/2$ 개의 비교에 대하여 자료산입을 수행하였다. 만약  $p$ 가 2, 3, 4, 5인 경우 자료산입이 수행되는 비교의 수는 각 1, 3, 6, 10 이다.  $p \geq 4$ 인 경우 연결되지 않은 비교의 수가 연결되어있는 비교의 수보다 크기 때문에 자료산입으로 생성되는 정보에 대한 비율이 높다.  $p \geq 4$ 인 개방형 네트워크에서 자료산입을 통한 보정법을 수행할 경우 개방형 네트워크에 포함된 기존의 근거를 비이상적으로 무력화시킬 수 있기 때문에 본 방법을 적용하는 것을 권장하지 않는다. 또한  $p$ 가 4 미만이라고 해도 연결되어 있는 비교의 수와 연결되지 않은 비교의 수가 비슷하므로 자료산입으로 생성되는 정보에 대한 비율이 여전히 높은 것으로 고려된다. 그러므로, 다중대체법 [45, 46]을 이용하여 자료산입을 통한 보정법을 적용한 민감도 분석 결과를 안정화하였다.

네트워크 메타분석에서 군-단위 (arm-level) 또는 연구-단위 (study-level)의 데이터 형식이 고려되지만, 본 연구에서는 연구-단위 데이터를 이용하여 방법을 설명하였다. 필요하다면, 공유 모수 모형(shared parameter model) [51]을 사용하여 여러 데이터 형식으로 구성된 데이터에 적용할 수 있도록 방법을 확장시킬 수 있다.

각 자료산입을 위해 생성한 데이터는 연구-단위의 치료효과 크기 ( $\hat{\theta}_{ijk}$ )와 이 값의 분산 ( $\widehat{var}(\hat{\theta}_{ijk})$ )으로 구성되어 있다. 우리는 각 비교에 대하여 생성한 효과 크기들의 분산값을 동일하게 설정하였다. [4.2.1.1. 자료산입 전략]에 따라 생성한 대체 데이터를 단순

메타분석으로 얻은 통합추정치(의 분산은 개방형 네트워크 메타분석으로 추정된 효과 크기의 분산과 같도록 하였다. 이 조건을 만족시키는데 있어서 분산들의 크기를 다양하게 설정하는 것은 불필요하다. 같은 맥락에서, 각 비교의 연구의 수에 대하여  $l \cdot \widehat{\text{var}}(\hat{d}_{Ak} - \hat{d}_{Aj})$  ( $j \neq k \neq A$ )이  $t^2$ 보다 크다는 제한조건하에 임의로 선택하도록 하였다.  $l$ 과  $1/\widehat{\text{var}}(\hat{\theta}_{kBC}^*)$ 은 서로 상쇄관계(trade-off relationship)이다.

본 연구에서는 여러 일관성 가정 검정 방법 중 관계없는 상대적 효과 평균을 통한 비일관성 모형[12]을 이용한 방법을 이용하였다. 민감도 분석 방법 개발에서는 네트워크의 전반적인 일관성 가정을 검정하기 위하여 광역적 검정 방법이 필요했고, 다른 모형을 이용한 광역적 검정 방법의 경우 네트워크에 포함된 닫힌 루프의 수가 충분히 크지 않은 경우 일관성 가정 검정에 대한 신뢰도가 낮다고 제안[12]되었기 때문에 관계없는 상대적 효과 평균을 통한 비일관성 모형을 이용한 검정 방법을 사용하였다.

Smoking cessation 데이터 세트로 만든 개방형 네트워크에서 자료산입을 통한 보정법을 적용한 결과는 원래의 완전한 네트워크에서 얻은 값과 비슷해졌다. 이는 실제 완전한 네트워크 결과가 개방형 네트워크의 결과보다 보수적인 상황에서, 자료산입을 통해 생성된 가상의 완전한 네트워크가 실제 완전한 네트워크를 상당히 유사하게 시뮬레이션 되었다는 것을 보여준다. 개방형 네트워크 메타

분석 결과의 강건성 평가 측면에서는, 자료산입을 통한 보정법을 이용한 민감도 분석의 67%에서 치료법의 순위가 개방형 네트워크에서 얻은 치료의 순위와 일치할 것이라고 결론 내릴 수 있다. 24개의 RCT가 모두 포함된 네트워크에서 각 비교별 추정된 비일관성 크기의 절대값은 최소 0.17, 최대 1.7인 범위에 있었고 전체적으로 비일관성에 대한 근거는 없었다[21]. 개발한 민감도 분석에서는 모든 대비에 대해 비일관성 크기의 절대값을 최대 1.05로 동일하게 가정하였고, 이 값은 상기 범위의 중간에 위치해있었다.

Simulated data set을 이용한 예제를 수행하여 연구 간 이질성이 비교적 작은 개방형 네트워크의 분석 결과가 더욱 강건하다는 것을 보였다. 네트워크 메타분석에서 비일관성과 이질성은 어떤 효과 조정 변수의 차이와 같은 요인들로 인해 나타나기 때문에 서로 관련이 있는 것으로 알려진바 있는데[12], 이와 같은 현상이 개발한 자료산입을 통한 보정법에 반영되고 있다는 것을 예제를 통해 보여주었다.

세 개 이상의 치료법으로 구성된 개방형 네트워크에 적용할 수 있도록 자료산입을 통한 보정법을 일반화할 수 있다.  $\bar{\omega}_{jk}$  for  $j = T_1, \dots, T_{p-1}$  and  $k = T_2, \dots, T_p$  ( $j \neq k$ ) 를 사용한다. 이 경우 각  $\bar{\omega}_{jk}$  는 같은 절대적 크기만큼 조건 I 을 충족하는 방향으로 동시에 변화한다. 하지만, 적절한 근거가 있다면 비일관성의 크기를 각 비교에 따라 지정할 수 있다. 예를 들어, 24개의 RCT가 포함된 smoking

cessation data set 의 경우, 전체 네트워크의 비일관성은 통계적으로 유의하지 않았지만, 각 비교별 비일관성 크기가 달랐다. 비교별 비일관성 크기의 다양성에 관한 사전 정보를 이용할 수 있는 경우 민감도 분석을 더욱 정교하게 수행할 수 있을 것이다.

개발한 자료산입을 통한 보정법과 유사한 개념 체계를 가진 방법으로 ‘임계 분석(threshold analysis)’이 Caldwell et al.(2016) [52]과 Phillippo et al.(2018, 2019) [53, 54]에 의해 제안된 바 있다[52-54]. 해당 방법은 네트워크 메타 분석에서 얻은 치료법에 대한 권장 사항에 대한 신뢰도를 평가하기 위한 민감도 분석 방법이다. 각 연구에서 추출한 효과크기(또는 네트워크 내에 추정된 비교-단위 효과 크기) 값의 위치를 변경시켜가며 분석을 수행하였을 때 최적의 치료법이 기존과 동일하게 도출되는지 또는 변경되는지 관찰하고, 최적의 치료법이 바뀌는 순간의 데이터 값, 즉, 임계값을 도출하여 결과의 신뢰도를 설명한다. 분석을 통해 얻은 임계값의 하한(lower)과 상한(upper)이 각 연구 또는 각 대조에 대한 기존 효과 크기의 95% 신뢰 구간 혹은 신용 구간 내에 있는 경우, 표본오차에 대한 혹은 있음직한 잠재적 비뚤림(bias)로 인하여 최적의 치료 권고사항이 변경될 가능성이 있다고 해석한다. 따라서, 임계 분석을 수행하여, 표본오차 혹은 있음직한 그러나 원인을 알 수 없는 잠재적 비뚤림에 대한 개방형 네트워크 메타분석의 신뢰성에 대한 평가가 가능하다. Smoking cessation data set 의 개방형 네트워크에 위 방법을 대조-단위(contrast-level)에서 적용하였다

(부록 B.4 참고). 임계 분석의 결과로부터 GC 보다는 IC가 최적의 치료일 수 있다는 가능성을 제시할 수 있었다. 분석에서 GC와 NI의 비교연구의 개수가 한 개였기 때문에, 이들 간 상대적 효과 크기의 신용구간(credible interval)이 넓었다. 결과적으로 현재 관찰된 효과 크기가 이 값의 신용구간 내에서 변화가 있다면 최적의 금연 프로그램이 변경될 수 있다. 본 연구에서 개발한 자료산입을 통한 보정법에서는 활성 치료법들의 사이의 가상의 직접비교 데이터를 분석에 통합하여 치료법들의 순위의 잠재적 변화를 시뮬레이션하였다. 직접비교 추정치는 관측된 간접비교 추정치로부터 허용 가능한 수준의 비일관성 범위 내에서 자료산입을 위해 생성되었다. 따라서 GC의 순위는 관측된 간접비교 근거와 약간의 비일관성을 포함하는 직접비교 데이터를 추가하는 것에 영향을 받지 않았으나 IC와 NI의 순위는 뒤바뀔 가능성이 있었다.

본 연구는 검증할 수 없는 일관성 가정 하에서 수행된 개방형 네트워크 메타 분석 결과의 강건성 평가에 초점을 맞췄다. 개발한 방법은 네트워크 내에 연결되지 않은 비교 수가 연결된 비교 수보다 적은 일부 열린 루프를 포함하는 네트워크에도 적용될 수 있다. 특히 네트워크에서 연결되지 않은 특정 치료군 간 비교의 비율이 상대적으로 작은 경우, 자료산입을 통한 보정법은 네트워크 메타분석에서 얻은 결과가 열린 루프의 알 수 없는 정보의 불확실성에 대하여 얼마나 견고할 수 있는지를 평가하는 도구로서 유용할 수 있다.



개별 환자 데이터를 사용할 수 있는 경우 population-adjusted indirect comparisons 방법을 적용하여 연구 간 환자 특성의 불균형을 개선할 수 있다[25-27]. 그러나, 만약 개별 환자 데이터의 환자 특성이 요약 데이터의 환자 특성과 유사하지 않은 경우 분석 결과에 비뚤림이 있을 수 있다. 이러한 경우에도 그 결과의 신뢰성을 평가하기 위하여 개발한 자료산입을 통한 보정법을 활용한 민감도 분석을 수행 할 수 있다.

## 제5장. 결론

서로 비교된 바 없었던 제2형 당뇨병 환자에게 인슐린 병합요법으로 투약하는 치료제들, DPP4i, GLP-1RA, SGLT2i, TZD 간 비교 효과 평가를 위해 교란요인 보정을 적용한 개방형 네트워크 메타분석을 수행하였다. 공통대조군인 Control군에서 관찰되는 결과 값들이 연구 간 상이함을 보였고 해당 값을 포함한 여러 교란요인들의 보정이 분석 결론에 미칠 수 있는 영향을 파악하여 제시하였다. 보정 전과 후 치료제들의 효과 순위는 동일하였고, 보정 전 과대 추정되었던 DPP4i의 효과 크기가 보정을 통해 감소하고 과소 추정되었던 GLP-1RA의 효과크기가 증가하여, 두 치료제들 간 효과 차이가 통계적으로 유의한 수준으로 나타났다. 본 분석을 통해 이질성이 보이는 RCT들로 구성된 개방형 네트워크에 대하여 일관성 가정에 근거를 통합하였을 때, 각 치료제들 간에 불분명하게 나타났던 치료효과 차이는 각 연구들에서 상이한 기저 위험 수준을 동일하게 보정함으로써 명확히 구분될 수 있었음을 확인 하였다. 또한, 검증이 불가능한 개방형 네트워크에서 전이성 가정을 최대한 만족하도록 하여 잠재적인 일관성 가정 위배를 최소화하는 접근의 중요성을 제시하였다.

네트워크 메타분석 접근을 사용하여 출판된 기존 연구 문헌들 중 개방형 네트워크 메타분석의 빈도 및 해당 연구들의 일관성에 대한 접근 방법과 결론의 한계점 등에 대한 경험적 근거 파악을 위

해 체계적 문헌 고찰을 수행하였다. 약 1년간 출판된 113개의 개방형 네트워크 메타분석 문헌을 검토한 결과, 네트워크 메타회귀분석 등의 구체적인 보정 방법을 적용하여 분석 결과에 반영한 연구는 단 10개에 지나지 않음을 확인하였다. 나머지 연구 중 96개의 연구에서는 일관성에 대한 가정 평가를 수행하지 않았거나(33개), 연구 특성의 정성적 비교만을 수행하여 일관성을 가정한 경우(57개), 또는 정성적 비교를 통해 비일관성이 예상됨에도 자료 수 부족으로 교란요인 보정 방법 고려가 불가능하였던 경우(6개)였으나, 모두 일관성 가정하에서 자료 통합을 수행하였다. 이처럼 출판된 대다수의 개방형 네트워크 메타분석들은 확신할 수 없는 일관성 가정하에 간접비교결과에만 기반하여 통합된 결과가 제시되고 있었음을 확인할 수 있었다. 본 체계적 문헌고찰을 통해 이들 연구에 실제적으로 직접비교 연구 자료가 포함될 수 있었다면 기존 결론이 유지될 수 있을 것인가에는 심각한 불확실성이 존재하며 이러한 불확실성이 최종적인 결론 및 의료 의사결정에 반영에 될 수 있어야함에 대한 문제를 제기하였다.

최종적으로 근거 네트워크에 포함된 연구들 간 특성의 정성적 평가만을 통해 일관성이 충족될 수 있을 것임을 가정하여 근거 통합이 수행되는 개방형 네트워크 메타분석에서 얻어지는 결과의 신뢰성 평가를 위한 방법론을 제시하였다. 개방형 네트워크 가상의 직접비교 연구 자료를 생성하여 산입한 후 통합하여 기존 결과를 보정하고 보정 후 기존 결론의 유지될 수 있을 것인지에

대한 확실성을 평가하는 자료산입을 통한 보정법을 개발하였다. 생성되는 가상의 직접 비교 자료는 주어진 간접비교 결과와 일정 크기의 비일관성이 존재하되 통계적으로 허용될 수 있는 수준이 되도록 하였고, 직접비교 결과는 간접비교 결과보다 보수적인 효과크기를 갖도록 조건을 설정하였다. 제안한 방법론을 통해 비일관성 크기의 허용범위를 찾고, 허용 범위 내에서 비일관성 증가에 따른 자료산입을 통해 보정 적용 결과를 도출한 후 기존 결과의 유지 정도를 평가하여 기존 결과의 신뢰성을 평가하였다. 예시로 주어진 완전한 네트워크에서 대안치료법들 간 RCT를 제외시켜 개방형 네트워크를 만들고 자료산입을 통한 보정법을 적용하였을 때의 결과와 본래의 완전한 네트워크 결과를 비교하여 본 방법의 타당성을 검증하였고, simulated data set을 이용한 예제를 수행하여 연구 간 이질성이 작을 수록 개방형 네트워크의 분석 결과의 신뢰성이 증가함을 보였다. 앞서 수행된 체계적 문헌 고찰 연구에서 파악된 중재 수가 4개 이하인 40개의 개방형 네트워크에 대하여 자료산입을 통한 보정법을 적용하여 보고된 각 결과의 신뢰성을 평가해 볼 수 있을 것이다.

의료기술평가 과정에서 네트워크 메타분석을 통해 제시된 근거에 기반하여 의사결정을 이루기 위해 참고하도록 여러 나라의 의료기술평가기관에서 제시한 지침에서는 네트워크 메타분석을 수행할 때 일관성에 대한 탐색 및 보고를 하도록 권고하고 있으며, 개방형 네트워크에 기반한 분석을 수행하여 결과를 제공할 때에는 교란요

인을 탐색하고 보정을 통하여 일관성 확보를 최대한 소명하여야 한다. 그러나 의료기술평가 현장에서 수행되는 많은 비교효과연구에서는 교란요인 보정을 수행하기에는 분석에 포함된 자료 수가 충분하지 않은 경우가 많으며 개별환자자료를 확보하는 것은 현실적으로 더욱 어렵다. 실제로 많은 연구들은 정성적인 평가에 기반하여 포함된 연구들이 임상적으로 매우 유사하다고 판단하고 이를 통해 일관성은 가정될 수 있다고 전제하여 분석을 수행하고 있으나, 이 경우 ‘통계적으로 허용될 수는 있으나 결과의 불확실성으로써 분석에 반영되어야 할’ 일정 수준의 비일관성은 실제로 존재할 수밖에 없고 이는 사실상 효과 추정 결과 및 추론에 반영되어야 함을 인지하여야 하며, 본 연구를 통해 개발된 방법론은 이와 같은 상황에 적용되어 결과의 신뢰성을 평가하고 보수적인 결론에 반영하기 위해 유용하게 활용될 수 있을 것이다. 더 나아가 각국의 의료기술평가 지침에서 자료산입을 통한 보정법 사용에 대한 가이드를 해준다면, 관심 중재 간 직접비교 연구가 부족한 경우에도 시기적절한 의사결정이 가능하게 될 것이다. 또한 향후 개편 예정에 있는 국내 심평원의 경제성 평가 지침 개정판에서도 네트워크 메타분석에 대한 제시안에 자료산입을 통한 보정법이 언급된다면, 확신할 수 없는 일관성 가정에 수행된 개방형 네트워크 메타분석으로부터 도출된 결과의 불확실성에 대응할 수 있을 것이다.

## 참고문헌

1. GS HJ: Cochrane handbook for systematic reviews of interventions version 5.1. 0. Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie 2011, 5(2):S38.
2. Bucher HC, Guyatt GH, Griffith LE, Walter SD: The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of clinical epidemiology 1997, 50(6):683-691.
3. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC: Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value in Health 2011, 14(4):417-428.
4. Caldwell DM, Ades A, Higgins J: Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ: British Medical Journal 2005, 331(7521):897.
5. Salanti G, Higgins JP, Ades A, Ioannidis JP: Evaluation of

- networks of randomized trials. *Statistical methods in medical research* 2008, 17(3):279–301.
6. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, Watanabe N, Nakagawa A, Omori IM, McGuire H: Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *The lancet* 2009, 373(9665):746–758.
  7. Wells G, Sultan S, Chen L, Coyle D: Indirect evidence: indirect treatment. *Comparisons in meta-analysis* Ottawa: Canadian Agency for Drugs and Technology in Health (CADTH) Health Technology Report 2009.
  8. Higgins JP, Green S: *Cochrane handbook for systematic reviews of interventions*, vol. 4: John Wiley & Sons; 2011.
  9. Cipriani A, Higgins JP, Geddes JR, Salanti G: Conceptual and technical challenges in network meta-analysis. *Annals of internal medicine* 2013, 159(2):130–137.
  10. Lu G, Ades A: Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in medicine* 2004, 23(20):3105–3124.
  11. Lumley T: Network meta-analysis for indirect treatment comparisons. *Statistics in medicine* 2002, 21(16):2313–2324.

12. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades A: NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials. NICE Decision Support Unit 2011.
13. Sutton A, Ades A, Cooper N, Abrams K: Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008, 26(9):753–767.
14. Yoon J–H, Min SH, Ahn CH, Cho YM, Hahn S: Comparison of non–insulin antidiabetic agents as an add–on drug to insulin therapy in type 2 diabetes: a network meta–analysis. *Scientific reports* 2018, 8(1):4095.
15. Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G: Effects of study precision and risk of bias in networks of interventions: a network meta–epidemiological study. *International journal of epidemiology* 2013, 42(4):1120–1131.
16. Lu G, Welton NJ, Higgins JP, White IR, Ades AE: Linear inference for mixed treatment comparison meta-analysis: a two-stage approach. *Research Synthesis Methods* 2011, 2(1):43–60.
17. Salanti G: Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many



- names, many benefits, many concerns for the next generation evidence synthesis tool. *Research synthesis methods* 2012, 3(2):80–97.
18. Jansen JP, Naci H: Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC medicine* 2013, 11(1):159.
  19. Lu G, Ades A: Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006, 101(474):447–459.
  20. Caldwell DM, Welton NJ, Ades A: Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *Journal of clinical epidemiology* 2010, 63(8):875–882.
  21. Dias S, Welton N, Caldwell D, Ades A: Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine* 2010, 29(7-8):932–944.
  22. Higgins J, Jackson D, Barrett J, Lu G, Ades A, White I: Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods* 2012, 3(2):98–110.

23. White IR, Barrett JK, Jackson D, Higgins J: Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* 2012, 3(2):111–125.
24. Drugs CAf, Health Ti: Guidelines for the economic evaluation of health technologies: Canada. In.: Ottawa: Canadian Agency for Drugs and Technologies in Health; 2017.
25. Cooper NJ, Sutton AJ, Achana F, Welton NJ: RFP Topic: Use of Network Meta-analysis to Inform Clinical Parameters in Economic Evaluations. Report for the CADTH 2015.
26. NioHaCE N: Guide to the methods of technology appraisal 2013. London: NICE 2013.
27. Hutton B, Catala-Lopez F, Moher D: The PRISMA statement extension for systematic reviews incorporating network meta-analysis: PRISMA-NMA. *Med Clin (Barc)* 2016, 147(6):262–266.
28. Committee PBA: Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Pharmaceutical Benefits Advisory Committee 2016.

29. Caro JJ, Ishak KJ: No head-to-head trial? Simulate the missing arms. *Pharmacoeconomics* 2010, 28(10):957–967.
30. Phillippo D, Ades T, Dias S, Palmer S, Abrams KR, Welton N: NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE. 2016.
31. Signorovitch JE, Wu EQ, Andrew PY, Gerrits CM, Kantor E, Bao Y, Gupta SR, Mulani PM: Comparative effectiveness without head-to-head trials. *Pharmacoeconomics* 2010, 28(10):935–945.
32. 경제성평가지침 및 자료작성 요령. 건강보험심사평가원 2011.
33. Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G: Characteristics of networks of interventions: a description of a database of 186 published networks. *PloS one* 2014, 9(1).
34. Phillippo DM, Dias S, Elstada A, Ades A, Welton NJ: Population Adjustment Methods for Indirect Comparisons: A Review of National Institute for Health and Care Excellence Technology Appraisals. *International journal of technology assessment in health care* 2019:1–8.

35. Salanti G, Marinho V, Higgins JP: A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *Journal of clinical epidemiology* 2009, 62(8):857–864.
36. Cooper NJ, Sutton AJ, Morris D, Ades A, Welton NJ: Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in medicine* 2009, 28(14):1861–1881.
37. Achana FA, Cooper NJ, Dias S, Lu G, Rice SJ, Kendrick D, Sutton AJ: Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Statistics in medicine* 2013, 32(5):752–771.
38. Higgins JP, Whitehead A: Borrowing strength from external trials in a meta-analysis. *Statistics in medicine* 1996, 15(24):2733–2749.
39. Song F, Altman DG, Glenny A-M, Deeks JJ: Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *Bmj* 2003, 326(7387):472.

40. Richter JE, Kumar A, Lipka S, Miladinovic B, Velanovich VJG: Efficacy of laparoscopic Nissen fundoplication vs transoral incisionless fundoplication or proton pump inhibitors in patients with gastroesophageal reflux disease: a systematic review and network meta-analysis. 2018, 154(5):1298–1308. e1297.
41. Rong LQ, Rahouma M, Abouarab A, Di Franco A, Calautti NM, Fitzgerald MM, Arisha MJ, Ibrahim DA, Girardi LN, Pryor KOJJoc et al: Intravenous and inhaled milrinone in adult cardiac surgery patients: A pairwise and network meta-analysis. 2019, 33(3):663–673.
42. Armoiry X, Tsertsvadze A, Connock M, Royle P, Melendez-Torres G, Souquet P-J, Clarke AJPo: Comparative efficacy and safety of licensed treatments for previously treated non-small cell lung cancer: A systematic review and network meta-analysis. 2018, 13(7):e0199575.
43. Ashaye AO, Khankhel Z, Xu Y, Fahrbach K, Mokgokong R, Orme ME, Lang K, Cappelleri JC, Mamolo CJFO: A comparative evaluation of gemtuzumab ozogamicin+ daunorubicin-cytarabine and other treatments for newly diagnosed acute myeloid leukemia. 2018, 15(6):663–681.

44. Kim WH, Hur M, Park SK, Jung D, Kang P, Yoo S, Bahk JHJA: Pharmacological interventions for protecting renal function after cardiac surgery: a Bayesian network meta-analysis of comparative effectiveness. 2018, 73(8):1019–1031.
45. Rubin DB: Multiple imputation after 18+ years. Journal of the American statistical Association 1996, 91(434):473–489.
46. Rubin DB: Multiple imputation for nonresponse in surveys, vol. 81: John Wiley & Sons; 2004.
47. Sturtz S, Ligges U, Gelman A: R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical software 2005, 12(3):1–16.
48. Team RC: R: a language and environment for statistical computing. Vienna, Austria; 2015. In.; 2015.
49. Hasselblad V: Meta-analysis of multitreatment studies. Medical Decision Making 1998, 18(1):37–43.
50. Higgins JP, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. Bmj 2003, 327(7414):557–560.
51. Dias S, Welton NJ, Sutton AJ, Ades A: NICE DSU technical support document 2: a generalised linear modelling

- framework for pairwise and network meta-analysis of randomised controlled trials. 2011.
52. Caldwell DM, Ades A, Dias S, Watkins S, Li T, Taske N, Naidoo B, Welton NJ: A threshold analysis assessed the credibility of conclusions from network meta-analysis. *Journal of clinical epidemiology* 2016, 80:68–76.
  53. Phillippo DM, Dias S, Ades A, Didelez V, Welton NJ: Sensitivity of treatment recommendations to bias in network meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2018, 181(3):843–867.
  54. Phillippo DM, Dias S, Welton NJ, Caldwell DM, Taske N, Ades A: Threshold analysis as an alternative to grade for assessing confidence in guideline recommendations based on network meta-analyses. *Annals of internal medicine* 2019, 170(8):538–546.

## 부록 A

### A.1. Summary of the studies included in the network meta-analysis

Study Source	Study duration of primary phase (weeks)	Insulin regimen	Study arms	Randomized participants, N	Age (years)	Proportion of male (%)	Baseline BMI (kg/m <sup>2</sup> )	Baseline HbA <sub>1c</sub> (%)	Duration of diabetes (years)	Baseline insulin dose (IU/day)
<b>Insulin plus DPP4 inhibitor</b>										
Barnett et al, 2012	24	stable	saxagliptin 5mg+insulin±meformin	304	57.2	40.0	32.6	8.7	11.8	53.6
			placebo+insulin±meformin	151	57.3	45.0	31.8	8.6	12.2	55.3
Fonseca et al, 2007	24	stable	vildagliptin 50mg bid+insulin	144	59.6	47.9	33.3	8.4	14.4	81.2
			placebo+insulin	152	58.9	54.6	32.9	8.4	14.9	81.9
Franc et al, 2015	12	NR	vildagliptin 50mg bid+insulin+meformin	31	59.4	NR	28.6	7.7	6.1	39.3
			placebo+insulin+meformin	31	59.4	NR	28.6	7.7	6.1	39.3
Hirose et al, 2015	12	stable	vildagliptin 50mg bid+insulin±meformin	78	58.5	70.5	25.3	8.1	12.8	20.7
			placebo+insulin±meformin	78	60.1	71.8	26.0	8.1	12.7	21.1
	16	stable	sitagliptin 50 mg qd+insulin	129	62.3	58.9	25.2	8.9	14.1	24.4



Kadowaki et al, 2013			placebo+insulin	137	60.2	58.4	25.2	8.9	14.0	24.3
Kaku et al, 2014	12	stable	alogliptin 25mg qd+insulin	90	62.9	55.6	23.9	8.4	15.3	22.6
			placebo+insulin	89	62.4	52.8	24.7	8.4	14.5	23.7
Kothny et al, 2013	24	stable	vildagliptin 50mg bid+insulin±metformin	228	59.3	47.8	28.9	8.8	12.9	39.9
			placebo+insulin±metformin	221	59.1	52.0	29.0	8.8	13.2	41.9
Mathieu et al, 2015	24	titration	sitagliptin 100mg per day+glargine ±metformin	329	59.3	45.9	31.9	8.7	13.2	37.3
			placebo+glargine±metformin	329	58.3	49.8	32.3	8.8	13.7	36.6
Ning et al, 2016	24	stable	vildagliptin 50mg bid+insulin±metformin	146	57.8	41.8	26.2	8.6	11.2	33.3
			placebo+insulin±metformin	147	58.4	44.9	26.0	8.7	11.4	31.7
Rosenstock et al, 2009	26	stable	alogliptin 25mg qd+insulin±metformin	129	55.9	34.0	32.3	9.3	13.4	55.0
			placebo+insulin±metformin	130	55.0	48.0	32.4	9.3	12.2	57.0
Sato et al, 2015	24	titration	sitagliptin 50mg or 100mg qd+insulin±OADs	25	66.0	64.0	24.5	7.9	19.0	31.8
			insulin±OADs	24	66.0	75.0	26.8	7.8	20.0	32.5
Shankar et al, 2016	24	stable	sitagliptin 100mg qd+insulin±metformin	234	58.6	55.6	25.9	8.7	11.0	34.5
			placebo+insulin±metformin	233	56.7	49.8	26.1	8.8	11.3	34.5
Vilsboll et al, 2010	24	stable	sitagliptin 100mg qd+insulin±metformin	322	58.3	49.0	31.0	8.7	13.0	44.2
			placebo+insulin±metformin	319	57.2	53.0	31.0	8.6	12.0	44.5
Yki-Jarvinen et al, 2013	24	stable	linagliptin 5mg qd+insulin±metformin±pioglitazone	631	59.7	52.1	30.8	8.3	NR	41.5

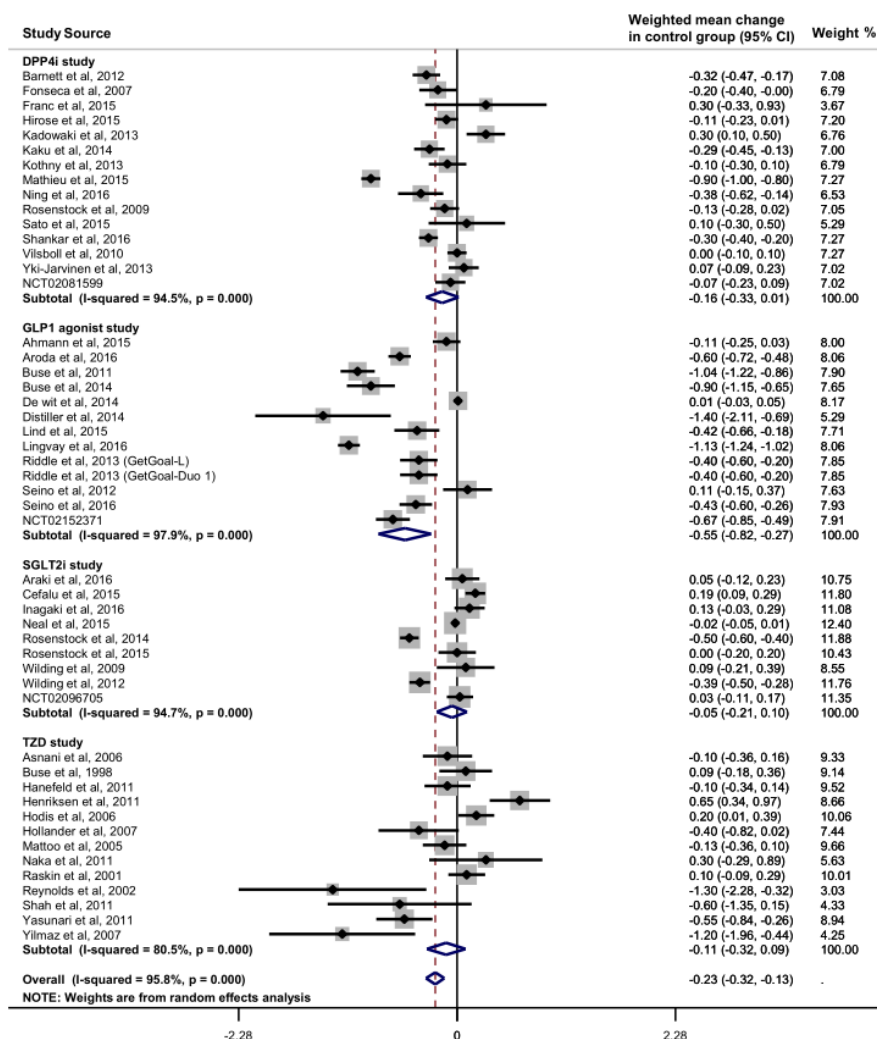
			placebo+insulin±metformin±pioglitazone	630	60.4	52.2	31.2	8.3	NR	40.1
NCT02081599	16	NR	teneligliptin 20mg qd+insulin	77	NR	76.6	NR	NR	NR	NR
			placebo+insulin	71	NR	74.6	NR	NR	NR	NR
Insulin plus GLP-1 receptor agonist										
Ahmann et al, 2015	26	stable	liraglutide 0.6-1.8mg qd+insulin±metformin	225	59.3	53.3	32.3	8.2	12.1	40.5
			placebo+insulin±metformin	225	57.5	60.4	32.2	8.3	12.1	40.5
Aroda et al, 2016	30	titration	fixed combination (lixisenatide+glargine) ±metformin	367	59.6	45.0	31.3	8.1	12.0	35.0
			glargine±metformin	369	60.3	48.5	31.0	8.1	12.1	35.2
Buse et al, 2011	30	titration	exenatide 10µg bid+insulin±metformin±pioglitazone	137	59.0	51.0	33.8	8.3	12.0	49.5
			placebo+insulin±metformin±pioglitazone	122	59.0	64.0	33.1	8.5	12.0	47.4
Buse et al, 2014	26	titration	liraglutide 0.6-1.8mg qd+degludec+metformin±SU/glinides	199	57.0	56.0	33.6	8.7	10.0	29.0
			plcebo+degludec+metformin±SU/glinides	199	58.0	53.0	33.8	8.8	11.0	29.0
de Wit et al, 2014	26	titration	liraglutide 0.6-1.8mg qd+insulin±metformin±SU	26	57.0	61.5	34.0	7.2	8.3	54.0
			insulin±metformin±SU	24	59.0	62.5	32.0	7.5	7.6	50.0
Distiller et al, 2014	24	titration	exenatide 10µg bid+insulin+metformin	14	49.1	50.0	41.9	8.7	12.3	253.0
			insulin+metformin	14	54.3	50.0	40.2	9.2	12.4	237.0
Lind et al, 2015	24	titration	liraglutide 0.6-1.8mg qd+insulin±metformin	64	63.7	62.5	33.7	9.0	17.3	105.3

			placebo+insulin±metformin	60	63.5	66.7	33.5	8.9	17.0	105.7
Lingvay et al, 2016	26	titration	fixed combination (liraglutide 0.6-1.8mg qd+degludec)±metformin	278	58.4	51.4	31.7	8.4	11.64	31.0
			glargine±metformin	279	59.1	49.1	31.7	8.2	11.33	32.0
Riddle et al, 2013 (GetGoal-L)	24	stable	lixisenatide 10-20µg qd+insulin±metformin	328	57.0	45.0	31.9	8.4	12.5	54.0
			placebo+insulin±metformin	167	57.0	49.0	32.6	8.4	12.4	58.0
Riddle et al, 2013 (GetGoal-Duo 1)	24	titration	lixisenatide 10-20µg qd+glargine +metformin±TZD	223	56.0	49.0	32.0	7.6	9.6	43.4
			placebo+glargine+metformin±TZD	223	56.0	51.0	31.7	7.6	8.7	44.2
Seino et al, 2012	24	stable	lixisenatide 20µg qd+insulin±SU	154	58.7	44.8	25.4	8.5	13.7	24.9
			placebo+insulin±SU	157	58.0	51.0	25.2	8.5	14.1	24.1
Seino et al, 2016	16	stable	liraglutide 0.9mg qd+insulin	127	61.3	54.3	26.2	8.8	14.32	30.0
			placebo+insulin	130	59.8	57.7	25.2	8.8	14.69	29.0
NCT02152371	28	NR	dulaglutide 1.5 mg qw+glargine ±metformin	150	60.2	56.7	NR	NR	NR	40.7
			placebo+glargine ±metformin	150	60.6	58.7	NR	NR	NR	36.6
<b>Insulin plus SGLT2 inhibitor</b>										
Araki et al, 2016	16	stable	dapagliflozin 5mg qd+insulin±DPP4i	122	58.3	73.0	26.89	8.3	15.32	37.9
			placebo+insulin±DPP4i	60	57.6	66.7	26.12	8.5	14.24	40.6
Inagaki et al, 2016	16	stable	canagliflozin 100mg qd+insulin	76	59.7	57.9	26.88	8.9	15.18	31.1
			placebo+insulin	70	56.1	70.0	25.99	8.9	12.34	28.1
Neal et al, 2015	18	stable	canagliflozin 300mg qd+insulin±OADs	690	63.0 <sup>a</sup>	65.0	33.3	8.3	16.3	60.0 <sup>a</sup>

			placebo+insulin±OADs	690	63.0 <sup>a</sup>	66.0	33.1	8.3	16.0	58.0 <sup>a</sup>
Rosenstock et al, 2014	18	stable	empagliflozin 25mg qd+insulin±metformin	189	58.0	44.0	35.0	8.3	NR	92.9
			placebo+insulin±metformin	188	55.3	40.0	34.7	8.3	NR	93.1
Rosenstock et al, 2015	18	stable	empagliflozin 25mg qd+insulin±metformin±SU	155	59.9	60.0	32.7	8.3	NR	48.4
			placebo+insulin±metformin±SU	170	58.1	53.0	31.8	8.2	NR	47.8
Wilding et al, 2009	12	stable	dapagliflozin 10mg qd+insulin±metformin±TZD	24	55.7	54.2	35.5	8.4	11.8	93.0
			placebo+insulin±metformin±TZD	23	58.4	69.6	34.8	8.4	13.8	80.0
Wilding et al, 2012	24	stable	dapagliflozin 10mg qd+insulin±OADs	194	59.3	44.8	33.4	8.6	14.2	78.0
			placebo+insulin±OADs	193	58.8	49.5	33.1	8.5	13.5	73.7
NCT02096705	24	NR	dapagliflozin 10mg qd+insulin	139	56.5	47.5	NR	NR	NR	NR
			placebo+insulin	133	58.6	48.1	NR	NR	NR	NR
Insulin plus thiazolidinedione										
Asnani et al, 2006	16	titration	pioglitazone 30mg qd+insulin±OADs	8	59.0	NR	NR	10.0	17.0	NR
			placebo+insulin±OADs	8	57.0	NR	NR	8.7	11.0	NR
Buse et al, 1998	26	titration	troglitazone 400mg qd+insulin	76	58.0	50.0	34.8	9.0	NR	NR
			placebo+insulin	71	57.0	49.0	34.5	9.0	NR	NR
Hanefeld et al, 2011	24	titration	pioglitazone 15mg bid+glargine +meformin	39	63.3	66.7	33.1	7.3	11.0	34.9
			placebo+glargine +meformin	42	64.2	54.8	31.8	7.4	12.3	36.6
Henriksen et al, 2011	26	stable	pioglitazone 45mg qd+insulin	102	60.1	69.0	33.2	8.7	13.8	78.4
			balaglitazone 20mg qd+insulin	97	60.5	55.0	34.1	8.5	14.7	80.2

			placebo+insulin	106	60.9	62.0	33.9	8.5	12.6	75.2
Hodis et al, 2006	24	titration	troglitazone 400mg qd+insulin	142	52.4	33.1	32.1	9.9	9.8	51.5
			placebo+insulin	134	52.6	32.1	31.1	9.7	9.7	54.5
Hollander et al, 2007	24	titration	rosiglitazone 2mg bid+insulin	189	52.6	48.1	33.7	9.0	13.0	73.5
			placebo+insulin	186	53.8	46.2	33.0	9.1	12.6	80.3
Mattoo et al, 2005	24	titration	pioglitazone 30mg qd+insulin±OADs	142	58.8	43.7	32.5	8.9	13.62	NR
			placebo+insulin±OADs	147	58.9	42.9	31.8	8.8	13.41	NR
Naka et al, 2011	24	NR	rosiglitazone 4mg qd+insulin	17	64.7	17.6	28.8	8.8	20.1	48.4
			insulin	14	67.3	28.6	29.0	8.8	17.1	47.1
Shah et al, 2011	12-16	titration	pioglitazone 45mg per day+insulin	12	58.0	84.0	36.7	7.6	NR	105.0
			placebo+insulin	13	58.0	84.0	36.7	7.8	NR	114.0
Yasunari et al, 2011	48	titration	pioglitazone 15mg bid+insulin+OADs	22	56.0	81.8	25.5	8.6	13.4	37.2
			insulin±OADs	26	57.2	69.2	26.9	8.6	14.8	39.6
Yilmaz et al, 2007	24	titration	rosiglitazone 8mg qd+insulin	15	57.6	53.3	30.7	9.6	12.1	41.9
			insulin	19	61.5	36.8	28.2	8.7	17.9	42.7

## A.2. Weighted mean change in HbA1c from baseline in control group



The change in HbA1c (%) from baseline in the controls groups of the DPP4i, GLP-1RA, SGLT2i, and TZD studies analyzed using a random-effects model. The squares indicate an individual study's effects and the size of the squares corresponds to the study's weight in the meta-analysis, with the horizontal lines extending from the symbols representing 95%

CI. The diamonds indicate the pooled estimates. (For the 3-arm trial, the treatment effect from only 1 treatment arm was presented.)

A.3. Pairwise results of comparisons between antidiabetic agents as an add-on to pre-existing insulin therapy from network meta-analyses adjusted by study-level covariates

Unadjusted results for difference in mean change of HbA1c from baseline (95% credible interval), %					
Control	-0.63 (-0.78, -0.49)	-0.76(-0.92, -0.60)	-0.69(-0.87, -0.52)	-0.74(-0.94, -0.54)	
-	DPP4i	-0.13(-0.34, 0.09)	-0.06(-0.29, 0.17)	-0.11(-0.35, 0.14)	
-	-	GLP-1RA	0.07(-0.17, 0.31)	0.03(-0.23, 0.28)	
-	-	-	SGLT2i	-0.04(-0.31, 0.22)	
-	-	-	-	TZD	
Probability (%) of being the highest ranked group for HbA1c reduction					
Control	DPP4i	GLP-1RA	SGLT2i	TZD	
0	3.24	52.08	28.08	16.60	
Adjusted results for difference in mean change of HbA1c from baseline (95% credible interval), %					
Control	-0.54(-0.68, -0.39)	-0.84(-1.00, -0.69)	-0.66(-0.84, -0.48)	-0.73(-0.93, -0.52)	
-	DPP4i	-0.30(-0.52, -0.09)	-0.11(-0.36, 0.12)	-0.18(-0.44, 0.07)	
-	-	GLP-1RA	0.19(-0.06, 0.43)	0.11(-0.16, 0.40)	
-	-	-	SGLT2i	-0.07(-0.33, 0.20)	
-	-	-	-	TZD	
Probability (%) of being the highest ranked group for HbA1c reduction					
Control	DPP4i	GLP-1RA	SGLT2i	TZD	
0	0.13	77.14	4.21	18.52	



## 부록 B

### B.1. $\widehat{var}(\hat{\theta}_{kBC}^*)$ , $k=1, \dots, l$ 를 계산하기 위한 공식 유도 과정

자료산입을 위해 생성한 데이터는 B와 C를 비교한 RCT 데이터  $\hat{\theta}_{kBC}^*$ 와  $\widehat{var}(\hat{\theta}_{kBC}^*)$  ( $k=1, \dots, l$ ,  $l$ 은 가상의 B와 C 비교 RCT의 수이다)로 구성되어 있다. 이 값들의 통합 추정치의 분산 값이  $\frac{1}{\sum_{k=1}^l \frac{1}{\widehat{var}(\hat{\theta}_{kBC}^*) + \hat{\tau}_{BC}^2}}$  이다[1].  $\hat{\tau}_{BC}^2$ 는 이 값들의 이질성에 대한 정도를 나타낸다. [4.2.1.1. 자료산입 전략]의 조건II와 III에서 위 통합 추정치의 분산 값은 개방형 네트워크 메타분석으로부터 얻은 효과 크기의 분산 값,  $\widehat{var}(\hat{d}_{AC} - \hat{d}_{AB})$ , 과 같고 또한,  $\hat{\tau}_{BC}^2$ 는 개방형 네트워크의 전체적인 이질성의 정도를 나타내는  $\hat{\tau}^2$ 와 같다고 가정하였다. 따라서, 다음 등식이 성립한다:

$$\widehat{var}(\hat{d}_{AC} - \hat{d}_{AB}) = \frac{1}{\sum_{k=1}^l \frac{1}{\widehat{var}(\hat{\theta}_{kBC}^*) + \hat{\tau}^2}} .$$

간결성을 위하여  $\widehat{var}(\hat{\theta}_{1BC}^*), \dots, \widehat{var}(\hat{\theta}_{lBC}^*)$  들은 모두 동일한 값을 갖는다고 설정하였다. 따라서 위 등식을 다음과 같이 표현할 수 있다:

$$\widehat{var}(\hat{d}_{AC} - \hat{d}_{AB}) = \frac{1}{\frac{l}{\widehat{var}(\hat{\theta}_{kBC}^*) + \hat{\tau}^2}} .$$

위 등식을 정리하여  $\widehat{var}(\hat{\theta}_{kBC}^*) = l \cdot \widehat{var}(\hat{d}_{AC} - \hat{d}_{AB}) - \hat{\tau}^2$ 을 얻었다

1. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in medicine*. 1991;10(11):1665–1677.

## B.2. R code : smoking cessation data set

```
library(R2WinBUGS)
```

```
# Calling WinBUGS code for consistency model and inconsistency model:  
# Adapt the WinBUGS code published in the supplement of:  
#Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades A.  
#NICE DSU Technical Support Document 4:  
#Inconsistency in networks of evidence based on randomised controlled trials.  
NICE Decision Support Unit. 2011.
```

```
modelfile<-"model_consistency.txt"  
modelfile2<-"model_inconsistency.txt"
```

```
smoking_cessation<-read.csv("smoking_cessation.csv",header=t) #loading t  
he smoking cessation data set  
##no intervention: treatment=1  
##self-help: treatment=2  
##individual counseling: treatment=3  
##group counseling: treatment=4
```

```
smoking_cessation
```

```
#> smoking_cessation
```

#	study	treatment	diff	std.err
#1	1	1	NA	NA
#2	1	2	-0.01596494	0.1699150
#3	2	1	NA	NA
#4	2	2	0.39350454	0.3265754
#5	3	1	NA	NA
#6	3	2	0.70294158	0.4413217
#7	4	1	NA	NA
#8	4	3	2.20228929	0.1430439

#9	5	1	NA	NA
#10	5	3	0.87035364	0.7910933
#11	6	1	NA	NA
#12	6	3	0.41564852	0.1557329
#13	7	1	NA	NA
#14	7	3	2.77968375	1.4698402
#15	8	1	NA	NA
#16	8	3	2.70539327	0.6251608
#17	9	1	NA	NA
#18	9	3	2.42518742	1.0422512
#19	10	1	NA	NA
#20	10	3	0.44361687	0.5219769
#21	11	1	NA	NA
#22	11	3	0.46459090	0.1417462
#23	12	1	NA	NA
#24	12	3	-0.15581043	0.3212230
#25	13	1	NA	NA
#26	13	3	-0.23997016	0.1736564
#27	14	1	NA	NA
#28	14	3	0.03895601	0.1873842
#29	15	1	NA	NA
#30	15	3	0.39041227	0.1680177
#31	16	1	NA	NA
#32	16	3	0.10633565	0.5955997
#33	17	1	NA	NA
#34	17	3	0.58339829	0.2983467
#35	18	1	NA	NA
#36	18	4	3.52251683	1.4969970
#37	19	2	NA	NA
#38	19	3	-0.15168459	0.4289753
#39	20	2	NA	NA
#40	20	4	1.04348631	0.4489795

```
#41    21          3          NA          NA
#42    21          4  0.68072466  0.4092394
#43    22          3          NA          NA
#44    22          4 -0.40546511  0.7139060
```

```
## network meta-analysis using the smoking cessation data set
```

```
t<-matrix(smoking_cessation$treatment,ncol=2,byrow=TRUE)
```

```
y<-matrix(smoking_cessation$diff,ncol=2,byrow=TRUE)
```

```
se<-matrix(smoking_cessation$std.err,ncol=2,byrow=TRUE)
```

```
na<-rep(2,nrow(t))
```

```
NS<-nrow(t)
```

```
NT<-length(levels(as.factor(smoking_cessation$treatment)))
```

```
inits<-function(){list(d=c(NA,0,0,0),tau=1)}
```

```
data<-list("t","y","se","na","NS","NT")
```

```
vars2keep<-list("d","tau","tau2","md","best")
```

```
output<-bugs(model.file=file.path(modelfile), data=data, inits = inits,
```

```
  parameters.to.save = vars2keep, n.chains=4, n.iter=25000, n.burnin=5
  000,
```

```
  n.thin=1, debug=TRUE, bugs.seed=1000, DIC=TRUE)
```

```
print(output,digits=4)
```

```
#> print(output,digits=4)
```

```
#Inference for Bugs model at "model_consistency.txt", fit using WinBU
GS,
```

```
# 4 chains, each with 25000 iterations (first 5000 discarded)
```

```
# n.sims = 80000 iterations saved
```

```
#          mean      sd    2.5%    25%    50%    75%  97.5%  Rhat
n.eff
#d[2]      0.4332 0.4166 -0.3763  0.1614  0.4277  0.6978  1.280 1.0011 170
00
```

```

#d[3]      0.7283 0.2393 0.2781 0.5693 0.7203 0.8787 1.225 1.0010 80
000
#d[4]      1.3757 0.5728 0.2691 0.9970 1.3670 1.7430 2.536 1.0010 69
000
#tau       0.7814 0.1875 0.4926 0.6481 0.7561 0.8857 1.220 1.0011 21
000
#tau2      0.6458 0.3315 0.2426 0.4200 0.5717 0.7844 1.489 1.0011 21
000
#md[1,2]   0.4332 0.4166 -0.3763 0.1614 0.4277 0.6978 1.280 1.0011 170
00
#md[1,3]   0.7283 0.2393 0.2781 0.5693 0.7203 0.8787 1.225 1.0010 80
000
#md[1,4]   1.3757 0.5728 0.2691 0.9970 1.3670 1.7430 2.536 1.0010 69
000
#md[2,3]   0.2951 0.4405 -0.5775 0.0132 0.2923 0.5767 1.171 1.0011 180
00
#md[2,4]   0.9425 0.6068 -0.2411 0.5456 0.9397 1.3340 2.155 1.0011 300
00
#md[3,4]   0.6474 0.5507 -0.4333 0.2846 0.6472 1.0030 1.741 1.0010 680
00
#best[1]   0.0000 0.0061 0.0000 0.0000 0.0000 0.0000 0.000 1.0513 800
00
#best[2]   0.0360 0.1862 0.0000 0.0000 0.0000 0.0000 1.000 1.0010 600
00
#best[3]   0.1002 0.3003 0.0000 0.0000 0.0000 0.0000 1.000 1.0010 380
00
#best[4]   0.8638 0.3430 0.0000 1.0000 1.0000 1.0000 1.000 1.0010 800
00
#deviance  21.9832 6.4900 10.9500 17.3500 21.3900 26.0300 36.230 1.0010 600
00
#
#For each parameter, n.eff is a crude measure of effective sample size,
#and Rhat is the potential scale reduction factor (at convergence, Rhat=
1).
#

```

```

#DIC info (using the rule,  $pD = Dbar - Dhat$ )
#pD = 16.9 and DIC = 38.9
#DIC is an estimate of expected predictive error (lower deviance is better).

##network meta-analysis using the derived open-loop network for the smoking cessation data set

open<-smoking_cessation[1:36,] ## creating an open-loop network

t<-matrix(open$treatment,ncol=2,byrow=TRUE)
y<-matrix(open$diff,ncol=2,byrow=TRUE)
se<-matrix(open$std.err,ncol=2,byrow=TRUE)
na<-rep(2,nrow(t))
NS<-nrow(t)
NT<-length(levels(as.factor(open$treatment)))

inits<-function(){list(d=c(NA,0,0,0),tau=1)}

data<-list("t","y","se","na","NS","NT")
vars2keep<-list("d","tau","tau2","md","best")

output<-bugs(model.file=file.path(modelfile), data=data, inits = inits,
  parameters.to.save = vars2keep, n.chains=4, n.iter=25000, n.burnin=5000,
  n.thin=1, debug=TRUE, bugs.seed=1000, DIC=TRUE)
print(output,digits=2)
#> print(output,digits=4)
#Inference for Bugs model at "model_consistency.txt", fit using WinBUGS,
# 4 chains, each with 25000 iterations (first 5000 discarded)
# n.sims = 80000 iterations saved

```

#	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
#d[2]	0.3317	0.5416	-0.7416	-0.0129	0.3272	0.6732	1.414	1.0010	470
00									
#d[3]	0.7213	0.2695	0.2092	0.5454	0.7122	0.8877	1.280	1.0010	72
000									
#d[4]	3.5240	1.7385	0.1406	2.3540	3.5160	4.6910	6.966	1.0011	15
000									
#tau	0.8499	0.2268	0.5133	0.6903	0.8148	0.9693	1.390	1.0010	80
000									
#tau2	0.7737	0.4531	0.2635	0.4765	0.6639	0.9395	1.932	1.0010	80
000									
#md[1,2]	0.3317	0.5416	-0.7416	-0.0129	0.3272	0.6732	1.414	1.0010	470
00									
#md[1,3]	0.7213	0.2695	0.2092	0.5454	0.7122	0.8877	1.280	1.0010	72
000									
#md[1,4]	3.5240	1.7385	0.1406	2.3540	3.5160	4.6910	6.966	1.0011	15
000									
#md[2,3]	0.3896	0.6049	-0.8025	0.0003	0.3849	0.7711	1.605	1.0010	400
00									
#md[2,4]	3.1923	1.8228	-0.3697	1.9630	3.1910	4.4173	6.781	1.0012	130
00									
#md[3,4]	2.8027	1.7564	-0.6355	1.6230	2.8000	3.9830	6.269	1.0011	170
00									
#best[1]	0.0000	0.0061	0.0000	0.0000	0.0000	0.0000	0.000	1.0513	800
00									
#best[2]	0.0193	0.1377	0.0000	0.0000	0.0000	0.0000	0.000	1.0010	800
00									
#best[3]	0.0416	0.1996	0.0000	0.0000	0.0000	0.0000	1.000	1.0010	800
00									
#best[4]	0.9390	0.2392	0.0000	1.0000	1.0000	1.0000	1.000	1.0010	800
00									
#deviance	15.8556	5.9804	5.9660	11.5900	15.2500	19.4500	29.200	1.0010	800
00									
#									

```

#For each parameter, n.eff is a crude measure of effective sample size,
#and Rhat is the potential scale reduction factor (at convergence, Rhat=
1).
#
#DIC info (using the rule,  $pD = Dbar - Dhat$ )
# $pD = 14.8$  and  $DIC = 30.6$ 
#DIC is an estimate of expected predictive error (lower deviance is better).

```

```

#####
#####
Application of our developed method to the data set corresponding to the
created open-loop network
#####
#####

```

```

##calculation of between-study and within-study variances

```

```

(tau<-0.8499)
(assumed_se23<-sqrt(5*(0.6049^2)-tau^2))
(assumed_se24<-sqrt(5*(1.8228^2)-tau^2))
(assumed_se34<-sqrt(5*(1.7564^2)-tau^2))

```

```

##defining a data set in which outputs from applying our method is saved

```

```

final<-data.frame(inconsistency= numeric(0),dev_con = numeric(0),dev_incon = numeric(0),
d12_sum= numeric(0),d13_sum = numeric(0),d14_sum = numeric(0),
d23_sum = numeric(0),d24_sum = numeric(0),d34_sum = numeric(0),
d12_var = numeric(0),d13_var = numeric(0),d14_var = numeric(0),
d23_var = numeric(0),d24_var = numeric(0),d34_var = numeric(0),
best1_sum= numeric(0),best2_sum= numeric(0),best3_sum = numeric(0),
best4_sum = numeric(0))

```



```

for (k in 1:100){
  dev1<-0;dev2<-0
  d12<-0;d13<-0;d14<-0;d23<-0;d24<-0;d34<-0
  sd12<-0;sd13<-0;sd14<-0;sd23<-0;sd24<-0;sd34<-0
  best1<-0;best2<-0;best3<-0;best4<-0

  for (j in 1:500){
    a<-rnorm(5,mean=0.3896-0.1*(k-1),sd=tau)
    b<-rnorm(5,mean=3.192-0.1*(k-1),sd=tau)
    c<-rnorm(5,mean=2.803-0.1*(k-1),sd=tau)
    a1<-0; b1<-0; c1<-0

    for(i in 1:5){
      a1[i]<-rnorm(1,a[i],sd=assumed_se23)
      b1[i]<-rnorm(1,b[i],sd=assumed_se24)
      c1[i]<-rnorm(1,c[i],sd=assumed_se34)
    }

    ##data imputation
    ## 2vs3
    data23_imputed<-data.frame(study=rep(19:23,each=2),treatment=c(rep(c("2",
"3"),5)),
    diff=c(NA,a1[1],NA,a1[2],NA,a1[3],NA,a1[4],NA,a1[5]),std.err=rep(c(NA,as
sumed_se23),5))
    ## 2vs4
    data24_imputed<-data.frame(study=rep(24:28,each=2),treatment=c(rep(c("2",
"4"),5)),
    diff=c(NA,b1[1],NA,b1[2],NA,b1[3],NA,b1[4],NA,b1[5]),std.err=rep(c(NA,a
ssumed_se24),5))
    ## 3vs4
    data34_imputed<-data.frame(study=rep(29:33,each=2),treatment=c(rep(c("3",

```

```
"4"),5)),
diff=c(NA,c1[1],NA,c1[2],NA,c1[3],NA,c1[4],NA,c1[5]),std.err=rep(c(NA,as
sumed_se34),5))
complete<-rbind(open,data23_imputed,data24_imputed,data34_imputed)
```

```
##network meta-analysis using a consistency model
```

```
t<-matrix(complete$treatment,ncol=2,byrow=TRUE)
```

```
y<-matrix(complete$diff,ncol=2,byrow=TRUE)
```

```
se<-matrix(complete$std.err,ncol=2,byrow=TRUE)
```

```
na<-rep(2,nrow(t))
```

```
NS<-nrow(t)
```

```
NT<-length(levels(as.factor(complete$treatment)))
```

```
inits<-function(){list(d=c(NA,0,0,0),sd=1)}
```

```
data<-list("t","y","se","na","NS","NT")
```

```
vars2keep<-list("d","tau","md","best")
```

```
output<-bugs(model.file=file.path(modelfile),data=data,inits = inits,
```

```
parameters.to.save = vars2keep,n.chains=3,n.iter=40000,n.burnin=10000,n.t
```

```
hin=20,
```

```
debug=FALSE,bugs.seed=1000,DIC=TRUE)
```

```
##network meta-analysis using an inconsistency model
```

```
inits2<-function(){list(d=structure(.Data=c(NA,NA,NA, 0,NA,NA, 0,0,NA,
0,0,0),.Dim=c(3,4)),sd=1)}
```

```
vars2keep2<-list("d","tau")
```

```
output2<-bugs(model.file=file.path(modelfile2),data=data,inits = inits2,
```

```
parameters.to.save = vars2keep2,n.chains=3,n.iter=40000,n.burnin=10000,n.
```

```
thin=20,
```

```
debug=FALSE,bugs.seed=1000,DIC=TRUE)
```

```
##abstraction of residual deviance from each model
```

```

output$mean$totresdev->dev1[j]
output2$mean$totresdev->dev2[j]
##abstraction of results from a consistency model
output$mean$md[1,1]->d12[j]
output$mean$md[1,2]->d13[j]
output$mean$md[1,3]->d14[j]
output$mean$md[2,1]->d23[j]
output$mean$md[2,2]->d24[j]
output$mean$md[3,1]->d34[j]

output$sd$md[1,1]->sd12[j]
output$sd$md[1,2]->sd13[j]
output$sd$md[1,3]->sd14[j]
output$sd$md[2,1]->sd23[j]
output$sd$md[2,2]->sd24[j]
output$sd$md[2,3]->sd34[j]

output$mean$best[1]->best1[j]
output$mean$best[2]->best2[j]
output$mean$best[3]->best3[j]
output$mean$best[4]->best4[j]
}

##combining the results from the 500 complete networks by Rubin's rule
e
final[k,1]<-0.1*(k-1)
final[k,2]<-mean(dev1)
final[k,3]<-mean(dev2)

final[k,4]<-mean(d12)
final[k,5]<-mean(d13)
final[k,6]<-mean(d14)

```

```

final[k,7]<-mean(d23)
final[k,8]<-mean(d24)
final[k,9]<-mean(d34)

final[k,10]<-mean(sd12^2)+var(d12)*(1+1/500)
final[k,11]<-mean(sd13^2)+var(d13)*(1+1/500)
final[k,12]<-mean(sd14^2)+var(d14)*(1+1/500)
final[k,13]<-mean(sd23^2)+var(d23)*(1+1/500)
final[k,14]<-mean(sd24^2)+var(d24)*(1+1/500)
final[k,15]<-mean(sd34^2)+var(d34)*(1+1/500)

final[k,16]<-mean(best1)
final[k,17]<-mean(best2)
final[k,18]<-mean(best3)
final[k,19]<-mean(best4)
if(mean(dev1)>mean(dev2)) break
}

```

### B.3. R code : simulated data set (a)

```
library(R2WinBUGS)
```

```
# Calling WinBUGS code for consistency model and inconsistency model:  
# Adapt the WinBUGS code published in the supplement of:  
#Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades A.  
#NICE DSU Technical Support Document 4:  
#Inconsistency in networks of evidence based on randomised controlled trials.  
NICE Decision Support Unit. 2011.
```

```
.
```

```
modelfile<-"model_consistency.txt" #calling the consistency model written  
in WinBUGS code
```

```
modelfile2<-"model_inconsistency.txt" #calling the inconsistency model written  
in WinBUGS code
```

```
s1<-read.csv("s1.csv",header=T) #loading the simulated data set (a)
```

```
##A: treatment=1
```

```
##B: treatment=2
```

```
##C: treatment=3
```

```
s1
```

```
#> s1
```

```
#   study treatment diff std.err  
#1      1          1   NA      NA  
#2      1          2   0.5      2  
#3      2          1   NA      NA  
#4      2          2   0.5      2  
#5      3          1   NA      NA  
#6      3          2   0.5      2  
#7      4          1   NA      NA
```

#8	4	2	0.5	2
#9	5	1	NA	NA
#10	5	2	0.5	2
#11	1	1	NA	NA
#12	1	3	1	2
#13	2	1	NA	NA
#14	2	3	1	2
#15	3	1	NA	NA
#16	3	3	1	2
#17	4	1	NA	NA
#18	4	3	1	2
#19	5	1	NA	NA
#20	5	3	1	2

```
## network meta-analysis using the data set
t<-matrix(s1$treatment,ncol=2,byrow=TRUE)
y<-matrix(s1$diff,ncol=2,byrow=TRUE)
se<-matrix(s1$std.err,ncol=2,byrow=TRUE)
na<-rep(2,nrow(t))
NS<-nrow(t)
NT<-length(levels(as.factor(s1$treatment)))

inits<-function(){list(d=c(NA,0,0),sd=1)}
data<-list("t","y","se","na","NS","NT")
vars2keep<-list("d","tau","sd","md","best")
output<-bugs(model.file=file.path(modelfile), data=data, inits = inits,
  parameters.to.save = vars2keep, n.chains=4, n.iter=25000, n.burnin=5
000,
  n.thin=1, debug=TRUE, bugs.seed=1000, DIC=TRUE)
print(output,digits=4)
#> print(output,digits=4)
#Inference for Bugs model at "model_consistency.txt", fit using WinBU
```

```

GS,
# 4 chains, each with 25000 iterations (first 5000 discarded)
# n.sims = 80000 iterations saved
#           mean      sd    2.5%    25%    50%    75%  97.5%
Rhat n.eff
#d[2]      0.5244 0.9856 -1.4260 -0.1282  0.5320  1.1820  2.442 1.0018 30
00
#d[3]      0.9924 0.9828 -0.8889  0.3337  0.9817  1.6360  2.969 1.0010 350
00
#tau       0.6846 0.5791  0.0262  0.2569  0.5416  0.9554  2.171 1.0026  1
500
#tau2      0.8040 1.4771  0.0007  0.0660  0.2933  0.9127  4.712 1.0026  1
500
#md[1,2]   0.5244 0.9856 -1.4260 -0.1282  0.5320  1.1820  2.442 1.0018 30
00
#md[1,3]   0.9924 0.9828 -0.8889  0.3337  0.9817  1.6360  2.969 1.0010 350
00
#md[2,3]   0.4680 1.3972 -2.2110 -0.4866  0.4476  1.4070  3.211 1.0014 64
00
#best[1]   0.0455 0.2085  0.0000  0.0000  0.0000  0.0000  1.000 1.0012 110
00
#best[2]   0.3532 0.4780  0.0000  0.0000  0.0000  1.0000  1.000 1.0027 15
00
#best[3]   0.6013 0.4896  0.0000  0.0000  1.0000  1.0000  1.000 1.0023 19
00
#deviance 35.2647 2.4392 32.4900 33.4900 34.6100 36.3800 41.640 1.0013 67
00
#
#For each parameter, n.eff is a crude measure of effective sample size,
#and Rhat is the potential scale reduction factor (at convergence, Rhat=
1).
#
#DIC info (using the rule, pD = Dbar-Dhat)
#pD = 3.0 and DIC = 38.3
#DIC is an estimate of expected predictive error (lower deviance is bett

```

er).

```
#####  
#####
```

Applying our developed method to the simulated data set

```
#####  
#####
```

```
##calculate between-study and within-study variances
```

```
(tau<-0.6846)
```

```
(assumed_se<-sqrt(5*(1.397^2)-tau^2))
```

```
##defin a data set in which outputs from applying our method is saved
```

```
final<-data.frame(inconsistency= numeric(0), dev_con = numeric(0), dev_i  
ncon = numeric(0),
```

```
d12_sum= numeric(0), d13_sum = numeric(0), d23_sum = numeric(0),
```

```
d12_var= numeric(0), d13_var = numeric(0), d23_var = numeric(0),
```

```
best1_sum= numeric(0), best2_sum= numeric(0), best3_sum = numeric(0))
```

```
for (k in 1:100){
```

```
dev1<-0;dev2<-0
```

```
d12<-0;d13<-0;d23<-0
```

```
sd12<-0;sd13<-0;sd23<-0
```

```
best1<-0;best2<-0;best3<-0
```

```
for (j in 1:500){
```

```
a<-rnorm(5,mean=0.5-0.2*(k-1),sd=tau)
```

```
a1<-0
```

```
for(i in 1:5){
```

```
a1[i]<-rnorm(1,a[i],sd=assumed_se)
```

```
}
```



```

##data imputation

##2vs3
data23_imputed<-data.frame(study=rep(12:16,each=2),treatment=c(rep(c("2",
"3"),5)),
diff=c(NA,a1[1],NA,a1[2],NA,a1[3],NA,a1[4],NA,a1[5]),
std.err=rep(c(NA,assumed_se),5))
complete<-rbind(s1,data23_imputed)

##network meta-analysis using a consistency model
t<-matrix(complete$treatment,ncol=2,byrow=TRUE)
y<-matrix(complete$diff,ncol=2,byrow=TRUE)
se<-matrix(complete$std.err,ncol=2,byrow=TRUE)
na<-rep(2,nrow(t))
NS<-nrow(t)
NT<-length(levels(as.factor(complete$treatment)))

inits<-function(){list(d=c(NA,0,0),sd=1)}
data<-list("t","y","se","na","NS","NT")
vars2keep<-list("d","tau","md","best")

output<-bugs(model.file=file.path(modelfile),data=data,inits = inits,
parameters.to.save = vars2keep,n.chains=3,n.iter=40000,n.burnin=10000,n.t
hin=20,
debug=FALSE,bugs.seed=1000,DIC=TRUE)

##network meta-analysis using an inconsistency model
inits2<-function(){list(d=structure(.Data=c(NA,NA,0,NA,0,0),.Dim=c(2,3)),s
d=1)}
vars2keep2<-list("d","tau")
output2<-bugs(model.file=file.path(modelfile2),data=data,inits = inits2,
parameters.to.save = vars2keep2,n.chains=3,n.iter=40000,n.burnin=10000,n.

```

```

thin=20,
debug=FALSE,bugs.seed=1000,DIC=TRUE)

##abstraction of residual deviance from result of each model
output$mean$totresdev->dev1[j]
output2$mean$totresdev->dev2[j]

##abstraction of results from a consistency model
output$mean$md[1,1]->d12[j]
output$mean$md[1,2]->d13[j]
output$mean$md[2,1]->d23[j]

output$sd$md[1,1]->sd12[j]
output$sd$md[1,2]->sd13[j]
output$sd$md[2,1]->sd23[j]

output$mean$best[1]->best1[j]
output$mean$best[2]->best2[j]
output$mean$best[3]->best3[j]
}

##combining the results from the 500 complete networks by Rubin's rule
e
final[k,1]<-0.2*(k-1)
final[k,2]<-mean(dev1)
final[k,3]<-mean(dev2)

final[k,4]<-mean(d12)
final[k,5]<-mean(d13)
final[k,6]<-mean(d14)

final[k,7]<-mean(sd12^2)+var(d12)*(1+1/500)

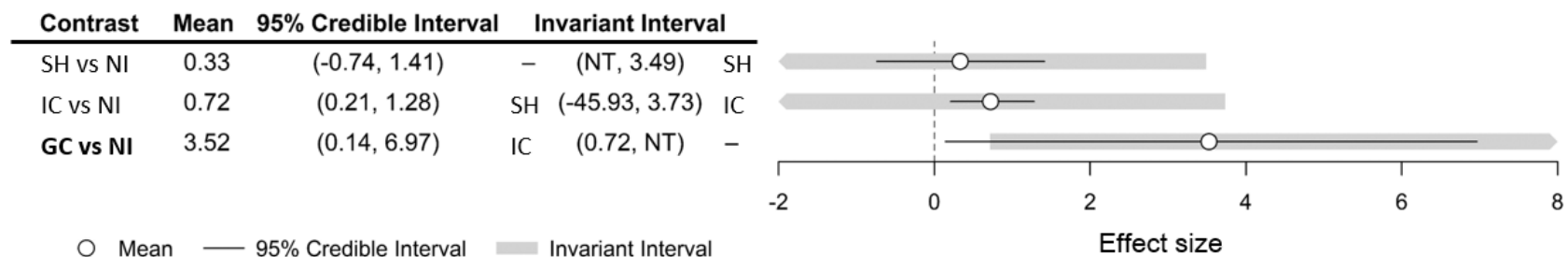
```

```
final[k,8]<-mean(sd13^2)+var(d13)*(1+1/500)
final[k,9]<-mean(sd14^2)+var(d14)*(1+1/500)

final[k,10]<-mean(best1)
final[k,11]<-mean(best2)
final[k,12]<-mean(best3)

if(mean(dev1)>mean(dev2)) break
}
```

#### B.4. Contrast-level forest plot with invariant intervals for the smoking cessation example



The base-case optimal treatment is GC. Bold letters in the table emphasize contrast estimates with short invariant intervals lying within the 95% credible interval. NI, no intervention; SH, self-help; IC, individual counseling; GC, group counseling.

## Abstract

# Development of a method for assessing the credibility of conclusions from open-loop network meta-analysis under unverifiable consistency assumption

Jeong-Hwa Yoon

Interdisciplinary Program in Medical Informatics

The Graduate School

Seoul National University

Network meta-analyses based on systematic reviews are often used to produce evidence for medical decision-making, such as deciding which of various treatment options is the best for a pre-defined population of patients. Specifically, network meta-analysis is a statistical method for integrating the data available from a network of multiple randomized controlled trials (RCTs) that involve multiple interventions compared directly, indirectly, or both. Since a network meta-analysis combining all information from RCTs on multiple interventions provides an

internally coherent set of estimates of the relative treatment effects between competing interventions, the comparability of the included trials should be ensured; that is, there should be no imbalance in the distribution of potential effect modifiers across the trials. The assumption of consistency across direct and indirect evidence should then be statistically checked. In an open-loop network, statistically detecting or checking inconsistency is impossible, but when a qualitative evaluation indicates that studies are comparable, researchers try to integrate the data into a network meta-analysis based only on indirect estimates under the consistency assumption.

The Canadian, British, and Australian Health Technology Assessment Guidelines require that the consistency between direct and indirect evidence within a network should be explored and reported when performing a network meta-analysis. The Canadian and Australian guidelines recommend that the results of a common comparator were used as a proxy for unmeasured patient characteristics which may be modifiers of intervention effect and therefore a potential source of heterogeneity and/or inconsistency. In particular, the Australian guideline suggests using a meta-regression analysis or methods for population-adjusted indirect comparisons to quantitatively identify factors

that may cause inconsistency.

In order to evaluate the comparative effectiveness among DPP4i, GLP-1RA, SGLT2i, and TZD as an adjunctive treatment in patients with poorly controlled type 2 diabetes mellitus on insulin therapy, which had never been compared with each other, an open loop network meta-analysis including 50 RCTs was performed. We identified as many potential confounders as possible and adjusted for their effect in the final analysis. In particular, the results of the common comparator were identified as an underlying risk indicator and an effect modifier. This exercise presented the importance of a covariate-adjustment approach that may neutralize potential inconsistency by ensuring transitivity in an open-loop network under unverifiable consistency assumption.

We systematically reviewed 113 open-loop network meta-analyses published during about 1 year to identify empirical evidence for their analytical limitations. In 10 studies there applied a covariate-adjustment method. In 96 studies there did not test for the essential assumptions (33 studies), or did qualitatively assess similarity between studies (57 studies), or did not apply covariate-adjustment due to a lack of data (6 studies). We confirmed that most of the published open-loop

network meta-analyses have presented incorporating results based on indirect comparisons under unverifiable consistency assumption. Through this systematic literature review, we raised the question of whether the existing conclusions could be maintained if direct comparison studies are included, and this uncertainty should also be reflected in the final conclusions and medical decision making.

Finally, we suggested a method of performing a sensitivity analysis through data imputation to assess the robustness of results from an open-loop network meta-analysis with an unknown degree of inconsistency. The method described herein involves data imputation for the disconnections in an open-loop network caused by the absence of RCTs for active treatments, producing a complete network. The imputed data simulate a situation that would allow for mixed treatment comparison, with a statistically acceptable extent of inconsistency. By comparing the agreement between the results obtained from the original open-loop network meta-analysis and the results after incorporating the imputed data, the robustness of the results of the original open-loop network meta-analysis can be quantified and assessed. We called this approach ‘adjustment through imputation’. To illustrate this method, we applied it to a real



data set and some simulated data sets. The adjustment through imputation applied to a hypothetically formed open-loop network by discarding all RCT data between active treatments from a real complete network showed that in 33% of the results from the analysis incorporating imputed data under acceptable inconsistency, the resulting treatment ranking would be inconsistent with the ranking obtained from the open-loop network. Through a simulation study, we demonstrated the sensitivity of results after data imputation for an open-loop network with different levels of within- and between-study variability. Our method will serve as a practical technique to assess the reliability of results from an open-loop network meta-analysis under the unverifiable consistency assumption.

**Keywords :** Network meta-analysis, consistency assumption, open-loop network, inconsistency, sensitivity analysis, data imputation

**Student Number :** 2014-30677